

DISSERTAÇÃO DE MESTRADO

Dinâmica da produção científica do Brasil

UNIVERSIDADE ESTADUAL DE MARINGÁ
MARINGÁ, 2008

DISSERTAÇÃO DE MESTRADO

ADRIANO ALVES BONACIN

DINÂMICA DA PRODUÇÃO CIENTÍFICA DO BRASIL

ORIENTADOR: Luís Carlos Malacarne

COORIENTADOR: Renio dos Santos Mendes

UNIVERSIDADE ESTADUAL DE MARINGÁ
MARINGÁ, 2008

Agradecimentos

Primeiramente à Deus, pois é quem tornou, torna e tornará tudo possível.

Ao professor Luís Carlos Malacarne, por me orientar neste mestrado e também ao professor Renio dos Santos Mendes, por grandiosas discussões, por seus finais de semana a mim dispensados e mais uma vez por ser exemplo de grande intelecto e de pessoa.

Aos grandes mestres da UEM, que mostraram o caminho a ser seguido honrando o compromisso que um dia firmaram.

À minha família, por sempre confiarem, pelo apoio, pelo aconchego e por entenderem minha ausência.

E, por fim, aos amigos e companheiros de república (os que dividiam o aluguel e os agregados), em quem busco exemplos de dedicação e sucesso e que muito motivam pelo fato de estarmos em busca do mesmo objetivo.

Ao professor Sérgio de Picoli Júnior, por seu auxílio quanto aos conceitos, pelas construtivas discussões e por ser um competente companheiro de caminhada, mesmo que distante.

À Tica, pelo auxílio às pressas com o Português.

*“Descobertas consistem em ver
o que todos viram e pensar
o que ninguém pensou.”*

A. Szent - Gyorgyi

Resumo

Neste trabalho, empregamos conceitos e métodos de física estatística para estudar a dinâmica de produção científica de centros de pesquisas brasileiros. Como consequência desta investigação, identificamos evidência de comportamento de invariância de escala e universalidade. Verificamos também que a produção científica brasileira vai na direção de um crescimento exponencial. Além disso, a distribuição da produção científica sugere uma forma lognormal que pode ser conectada com o modelo de Gibrat. Por outro lado, nossas análises indicam que as taxas de crescimento seguem uma distribuição exponencial (Laplace) e que o desvio padrão das taxas de crescimento decresce com a produção científica como uma lei de potência. Tais aspectos são bem reproduzidos por uma generalização do modelo de Gibrat. Esses resultados são notavelmente semelhantes aos observados na dinâmica das atividades econômicas e de pesquisas científicas. Tais comportamentos também são comparados com a dinâmica de produção científica dos Estados Unidos da América e da Europa.

Abstract

In this work, we employ concepts and methods of statistical physics to study the dynamics of scientific output of Brazilian research centers. As a consequence of this investigation, we identify evidence of scaling behavior and universality. We also verify that the Brazilian scientific output goes towards an exponential growth. Moreover, the distribution of the scientific output suggests a lognormal shape which may be connected with the Gibrat's model. However, our analysis indicates that the growth rate follows an exponential distribution (Laplace) and the standard deviation of the growth rate decreases with the scientific output like a power law. These aspects are well reproduced by a generalization of Gibrat model. Our findings are remarkably similar to those observed in the dynamics of economic and scientific research activities. These behaviors are also compared with the dynamics of scientific output from the United States of America and Europe.

Lista de Figuras

1.1	Em a), vemos a representação do decrescimento exponencial em escala linear e, em b), o mesmo decrescimento em um gráfico mono-log. Em c), o crescimento exponencial com escala linear é exibido, e em d), com escala mono-log.	6
1.2	Crescimento do Banco de Dados GenBank. Na escala à esquerda, é apresentado o número de seqüências catalogadas (em milhões), e à direita o número de pares de bases (em bilhões). Imagem retirada do <i>site</i> do NCBI [31].	7
1.3	Crescimento do número de pacotes do sistema operacional FreeBSD. Figura retirada da ref. [32].	8
1.4	Crescimento dos idiomas em que a bíblia foi traduzida e publicada, em apenas alguns livros, um testamento ou inteira. Figura retirada da ref. [33].	9
1.5	Crescimento da capacidade de processamento, associada ao número de transistores. Note que o gráfico é mono-log e por isso uma reta.	10
1.6	As curvas logísticas são representadas pela equação (1.8). Em preto, temos uma curva exponencial para ilustrar a tendência inicial do crescimento logístico, note que ela satisfaz a equação (1.9).	11
1.7	Estimativa do <i>US Census Bureau</i> para a população mundial, em bilhões, ao longo dos anos.	12
1.8	Relação entre proporção de insetos mortos e a dose aplicada.	12
1.9	Estimativa do número de casos de AIDS na África sub-Saariana, para adultos, homens e mulheres, com idade entre 15 e 49 anos. Os quadrados representam os homens e os triângulos as mulheres. Fonte UNAIDS[38]. . . .	13
1.10	Figura retirada e adaptada da referência [39]. Aqui, representa-se a razão entre os gastos (G) e a arrecadação (A) do governo norte-americano ao longo dos anos. Esse crescimento segue uma curva logística.	14
1.11	Figura retirada da referência [40]. A evolução da massa seca do fruto da manga <i>cv. Haden</i> segue uma curva logística.	14
1.12	Em a), mostramos a representação da equação (1.10) em escala linear. Os parâmetros estão indicados na figura. Em b), a escala logaritma é utilizada em ambos os eixos.	15
1.13	Figura retirada da referência [47]. A distribuição de tamanhos de agregados, n_s , segue uma função do tipo lei de potência com o aumento do número de células, S , representada por retas em gráficos log-log. Aqui $-\tau$ é o expoente da lei de potência.	16

1.14	Figura retirada e adaptada da referência [48]. Representa-se a probabilidade, $p(k)$, de cada vértice interagir diretamente com outros k vértices. Em a), o sistema observado é a <i>internet</i> , e a linha pontilhada é uma função $p(k) \propto k^{-2,1}$. Já em b), representa-se a rede das linhas de distribuição de energia elétrica e a linha pontilhada é uma função $p(k) \propto k^{-4}$	17
1.15	Figura retirada e adaptada da referência [49]. Distribuição de probabilidades para os blecautes ocorridos nos Estados Unidos da América.	18
1.16	Figura retirada e adaptada da referência [50]. A distribuição acumulada (veja a discussão correspondente a equação (3.7)) do número de vítimas para diversos tipos de ataques.	19
1.17	O gráfico, em escala linear, de uma distribuição gaussiana tem a forma de um sino, como vemos em a). Quando utilizamos escala logaritma no eixo das ordenadas obtemos uma parábola, como visto em b).	20
1.18	Histograma representando a distribuição de frequências da altura de seres humanos. Note que o valor médio é diferente para homens e mulheres.	20
1.19	Figura retirada e adaptada da referência [53]. A distribuição de frequência do número de crianças com a dada concentração de hemoglobina no sangue.	21
1.20	A distribuição lognormal em gráfico com escala linear em a) e em log-log em b), onde o gráfico tem a forma de uma parábola. Em vermelho, mostra-se a distribuição com $\mu = 1,0$ e $\sigma = 1,0$, já em azul, $\mu = 1,5$ e $\sigma = 1,0$ e finalmente em preto, $\mu = 1,5$ e $\sigma = 1,5$	22
1.21	Figura retirada e adaptada da referência [61]. A distribuição lognormal em futebol, em que $\ln(t_{out}/s)$ representa o logaritmo do tempo que fica parado um jogo de futebol devido a faltas, arremessos laterais, etc. A curva contínua representa uma parábola e os círculos são dados observacionais.	23
1.22	Histograma para dados aleatórios simulados seguindo uma distribuição de Laplace com desvio padrão unitário e média nula. Em a), exibimos o gráfico normal, em b), o gráfico do logaritmo da distribuição e, em c), o mono-log. A linha contínua representa a distribuição e os pontos o histograma.	24
2.1	Figura retirada e adaptada da referência [21], apresenta, em a), a distribuição de probabilidades condicional para pequenos e grandes PIB's. Nota-se aqui que embora apresentem a mesma forma funcional, as larguras das distribuições são diferentes. As linhas tracejada e contínua correspondem a equação (2.8). Em b), empregada a taxa de crescimento normalizada $r = (R - \mu)/\sigma$, vê-se as distribuições de probabilidades colapsando em uma única curva, ainda com a mesma forma. A linha tracejada corresponde a uma função a $\exp(-\sqrt{2} r)/\sqrt{2}$	28
2.2	Figura retirada da referência [23]. Nota-se que, após a normalização, dados relativos a diferentes campos colapsam em uma única curva.	28
2.3	Histograma obtido a partir da simulação do modelo de Gibrat com 2×10^4 partes e 10^5 iteradas. Em a), é exibido o gráfico usual da lognormal (linha contínua) e os pontos compõem o histograma dos dados simulados. Em b), o logaritmo da lognormal pelo logaritmo de z é representada por uma parábola (linha contínua) e os pontos dizem respeito à simulação.	30

- 2.4 Figura retirada da referência [13]. Realização típica do modelo definido pelas equações 2.10 e 2.11 com 1000 elementos de um sistema, todos com valor inicial de $F_0 = 1$, com $F \geq F_m = 1$. Fixou-se $b_0 = 0,01$ e variou-se os valores de b_1 e k . Em cada caso foi considerado 100 passos depois de um regime inicial. (a) Densidade de probabilidade $p(r)$, para $k = 1$ e $b_1 = 0$ (quadrados), $b_1 = 0,6$ (círculos) e $b_1 = 0,9$ (triângulos). (b) Densidade de probabilidade das taxas de crescimento normalizadas, $p(r)$, para $b_1 = 0,6$ e $k = 1$ (círculos), $k = 0,8$ (quadrados) e $k = 0,5$ (triângulos). Em (a) e (b) as curvas foram deslocadas na vertical para melhor visualização. (c) Desvio padrão das taxas de crescimento, $\sigma(F)$, para $b_1 = 0,6$ e $k = 0,7$ (quadrados), $k = 0,8$ (círculos) e $k = 0,9$ (triângulos). As linhas retas são ajustes lineares representando leis de potência com expoente $\beta \simeq 0,3$, $\beta \simeq 0,2$ e $\beta \simeq 0,1$, respectivamente. (d) Desvio padrão das taxas de crescimento, $\sigma(F)$, para $k = 0,8$ e $b_1 = 0$ (quadrados), $b_1 = 0,6$ (círculos) e $b_1 = 0,8$ (triângulos). As linhas retas representam leis de potência com expoente $\beta = 0,2$ 32
- 3.1 Figura retirada da referência [12]. Em a), vê-se a densidade de probabilidade $p(\log F)$ para os referidos periódicos (círculos), em que F é o fator de impacto. A linha sólida representa uma distribuição gaussiana na variável $\log F$ (portanto, uma lognormal em F). Em b), evidencia-se que a distribuição $p(F)$ tem um comportamento do tipo lei de potência para pequenos e grandes valores de F . A linha pontilhada a esquerda representa uma função $p(F) \propto F^{0,4}$, a pontilhada a direita uma função $p(F) \propto F^{-2,7}$ e a linha sólida $p(F) \propto F^{0,4}(1,3 + F^2)^{-1,55}$ 34
- 3.2 Figura retirada da referência [12]. Em a), tem-se as distribuições de probabilidades condicionais $p(R|F)$, para três grupos distintos, usando como critério o tamanho do fator de impacto, $0,1 < F < 0,2$ (quadrados), $0,3 < F < 0,7$ (círculos) e $1,3 < F < 2$ (triângulos). A mesma forma funcional é verificada, porém com diferentes larguras. Em b), o desvio padrão decresce com o valor do fator de impacto F , justificando as diferenças de largura em a). Em c), a normalização faz com que haja o colapso das distribuições de a) em uma única curva. 35
- 3.3 Figura retirada da referência [25]. Em a), ilustra-se a distribuição dos 100 maiores jornais dos EUA em 2003 (triângulos para baixo), 100 maiores revistas do RU em 2004 (triângulos para cima), 100 maiores revistas dos EUA (quadrados) e 100 maiores jornais do RU (círculos), com o expoente $\alpha \simeq 1,5$. Em b), em ambos os conjuntos de dados, $P_C(S)$ apresenta um comportamento assintótico do tipo lei de potência. 36
- 3.4 Figura retirada da referência [25]. Tem-se o desvio padrão das taxas de crescimento em função da circulação, em a), e a distribuição das taxas de crescimento normalizada para revistas com grande circulação, em b). . . . 37

3.5	Figura retirada e adaptada da referência [23]. Em a), um gráfico do logaritmo da distribuição de probabilidades do logaritmo dos gastos com Pesquisa e Desenvolvimento (P&D) nos EUA é apresentado. A linha sólida representa um ajuste gaussiano. Em b), as densidades de probabilidades para três grupos distintos usando como critério os valores de seus gastos com P&D, a mesma forma funcional é encontrada. Como em b) as larguras são diferentes, esperamos que o desvio padrão decresça com o aumento do tamanho dos gastos, verificado em c). Realizada a normalização, em d), as distribuições de probabilidade dos três grupos colapsam em uma única curva.	39
3.6	Figura retirada e adaptada da referência [23]. Em a), vemos o desvio padrão das taxas de crescimento para diferentes setores de diferentes países. A linha sólida indica uma lei de potência com expoente igual a $-0,25$. A idéia de universalidade é reforçada em b), onde a distribuição da densidade de probabilidade normalizada para os diferentes setores de diferentes tamanhos e de diferentes países colapsam em uma única curva. Em c), o colapso se dá na distribuição da densidade de probabilidade para universidades, para firmas e para países.	40
3.7	Figura retirada e adaptada da referência [9]. Como primeiro resultado tem-se uma distribuição de probabilidades bi-modal para o tamanho das publicações dos países, segundo a contagem fracionária. Os resultados obtidos com as contagens inteiras são análogos.	42
3.8	Figura retirada e adaptada da referência [9]. Vemos, em a), o decaimento do desvio padrão, σ , com o aumento do número de publicações, S . Em b), as distribuições de probabilidades normalizadas para os três diferentes grupos. Note o colapso dos dados em uma única curva com a forma de tenda (<i>tent-shape</i>).	43
3.9	Figura retirada e adaptada da referência [9]. Histograma do logaritmo do tamanho da publicação de: a) 408 institutos dos Estados Unidos e b) 508 institutos da União Européia, medidos com o sistema de contagem fracional, para o período de 11 anos (1991-2001).	44
3.10	Figura retirada e adaptada da referência [9]. As instituições européias e norte-americanas foram divididas em 10 grupos de acordo com o tamanho de suas publicações, S . Percebemos então, que o desvio padrão, $\sigma(S)$, decai segundo uma lei de potência com expoente $-0,3$ para os institutos norte-americano (círculos) e $0,39$ para os institutos europeus (quadrados).	45
3.11	Figura retirada e adaptada da referência [9]. Histograma do logaritmo do número de publicações de 2330 autores belgas ao longo de 22 anos (1980-2001) para o sistema de contagem fracionária a) e inteira b). As linhas cheias são ajustes gaussianos, assim como os previstos por Gibrat, para sistemas em que as taxas de crescimento são independentes e não correlacionadas com o tamanho do sistema, com distribuição de probabilidades lognormal.	46

3.12	Figura retirada e adaptada da referência [9]. Utilizando o sistema de contagem fracionário para as publicações dos autores belgas, em a), foram divididos em 10 grupos de acordo com o tamanho de suas publicações. Foi encontrado que o desvio padrão, σ , decai segundo uma lei de potência de expoente $-0,28$. Já em b), a partir da distribuição de probabilidades normalizadas para as taxas de crescimento, vê-se o colapso de três diferentes setores.	47
4.1	Ajuste exponencial para as quatro instituições científicas brasileiras mais produtivas (em número de artigos).	50
4.2	Embora optamos pelo ajuste exponencial, a produção científica de algumas instituições podem ser bem ajustadas por uma função logística. À esquerda o ajuste é feito com uma função logística e à direita com uma função exponencial, para as mesmas instituições.	51
4.3	Em a), vemos como os coeficientes A 's apresentam um decaimento conforme diminui a produtividade dos institutos (classificação em ordem decrescente) e, em b), fica evidenciado o crescimento com o aumento da produtividade, aqui para as 36 instituições mais produtivas. A linha sólida à esquerda corresponde a lei de potência: $A(n) = 92,94 n^{(-0,90)}$, em que n é um número que representa a classificação do instituto, à direita indica também uma lei de potência, $A(s) = 0,50 s^{0,89}$, com s representando a produtividade no ano de 2004.	51
4.4	Em a), notamos que os coeficientes B 's apresentam uma leve tendência de crescimento, quando consideramos todas as instituições de pesquisa. Em contraste, quando nos restringimos às instituições mais produtivas, é evidenciada uma constância, em média, dos B 's. Para as instituições mais produtivas, em b), temos os coeficientes B 's em função do logaritmo da produtividade total de cada instituto, S . E finalmente, em c), os coeficientes em função do logaritmo da produtividade no ano de 2004, $\log(s)$, para as instituições mais produtivas. A linha tracejada corresponde a $B(n) = 0,085 + 0,001n$, e as linhas sólidas marcam 0,12, o valor médio dos coeficientes B 's	52
4.5	Evolução da produtividade científica do Brasil (em artigos) em gráfico usual, em a), e em gráfico mono-log, em b). Os pontos representam a produtividade brasileira em cada ano até 2004. As linhas contínuas correspondem a função $s(t) = 258,63 \exp(0,12t)$	52
4.6	A distribuição de probabilidade da taxa de crescimento segue um padrão exponencial, como nos exemplos discutidos no capítulo anterior. Percebemos também que seu valor médio é positivo e, conseqüentemente, a tendência é que, em média, as instituições cresçam ao passar dos anos.	53

4.7	Em azul, a distribuição de probabilidades das taxas de crescimento das vinte maiores (primeiras) instituições (em número total de artigos publicados). Em verde, temos a distribuição das taxas de crescimento também de vinte instituições, desde a octagésima primeira até a centésima, segundo a mesma classificação. A forma da distribuição é mantida embora as larguras sejam diferentes, o que sugere um desvio padrão maior nas menores instituições. As curvas contínuas correspondem a ajustes usando a equação (2.8).	54
4.8	Com a normalização (uso de $r = (R - \mu)/\sigma$), a distribuição das taxas de crescimento colapsam numa única curva. A curva contínua representa a equação (2.8) com desvio padrão unitário, $\sigma = 1$ e média nula, $\mu = 0$	55
4.9	Percebemos que a distribuição dos tamanhos das instituições de pesquisa brasileiras é aproximadamente lognormal, visto que é bem ajustada por uma parábola. A janela utilizada, na variável $\log[s]$, foi de 0,6.	55
4.10	Há uma tendência de decrescimento do desvio padrão à medida que aumenta a produtividade dos institutos. Esse decrescimento segue uma lei de potência, com expoente $-0,15$	56
C.1	Página inicial do Web Of Science usada durante a pesquisa.	66

Sumário

Introdução	1
1 Leis de Crescimento e Distribuições de Probabilidade	4
1.1 Crescimento Exponencial	4
1.1.1 Exemplos	5
1.2 Crescimento Logístico	9
1.2.1 Exemplos	10
1.3 Distribuição do Tipo Lei de Potência	15
1.3.1 Exemplos	15
1.4 Distribuição Gaussiana	18
1.4.1 Exemplos	20
1.5 Distribuição Lognormal	21
1.5.1 Exemplos	21
1.6 Histogramas	22
2 Séries Temporais	25
2.1 Introdução às Séries Temporais	25
2.2 Análise das Flutuações	27
2.3 Modelo de Gibrat e generalização	29
3 Aplicações Recentes	33
3.1 Citações de Periódicos Científicos	33
3.2 Circulação de Revistas e Jornais	36
3.3 Pesquisas em Universidades	37
3.4 Produção Científica	41
4 Produção Científica do Brasil: Centros de Pesquisa	48
4.1 Introdução	48
4.2 Métodos Utilizados	48
4.3 Comportamento Exponencial	49
4.4 Distribuição das Taxas de Crescimento	53
4.5 Distribuição do Tamanho das Instituições	54
4.6 Desvio Padrão das Taxas de Crescimento	55
5 Conclusão	57
A Teorema Central do Limite	60

B	Distribuição Lognormal	63
C	Coleta dos Dados	65
D	Gráficos	67

Introdução

A socialização do conhecimento é algo crucial quando falamos em ciência, o que não se observa em outros setores, como o de tecnologia, em que o segredo muitas vezes é questão de sobrevivência em um mercado extremamente dinâmico e competitivo.

Ao longo dos anos, muitos periódicos de divulgação científica passaram a existir e com eles vasta quantidade de informações. Surgiram, também, instituições com o intuito de agrupar os vários dados bibliográficos, como é o caso do *Institute for Scientific Information* (ISI) [1], do *Elsevier* [2], entre outros [3, 4, 5]. Relacionados a esses há ainda muitas pesquisas investigando números referentes à produção científica. Mais do que isso, existe toda uma área do conhecimento direcionada a “estudar e tratar o planejamento, a implementação, a administração e a organização da informação em unidades de informação (das quais podem ser citadas as bibliotecas, centros de documentação e informação, sistemas de informação, entre outros)” [6], a biblioteconomia. Nesta dissertação, em particular, seguimos a linha de investigação relacionada às referências [7, 8, 9, 10, 11, 12, 13]. Mais precisamente, investigaremos alguns aspectos relacionados à dinâmica de crescimento das universidades e institutos de pesquisa brasileiros referenciados pelo seu respectivo número de publicações. Coletamos o número de artigos publicados por instituição ao longo dos anos e, com estas informações, analisamos aspectos estatísticos típicos deste tipo de dados tendo como base de dados o *Web of Science*, do ISI.

Um ponto marcante da produção científica é a interação entre seus elementos, onde a publicação de um trabalho pode estimular outros pesquisadores ou mesmo obrigá-los a mudar o foco de suas pesquisas. Essa situação faz com que esse sistema (produção científica) possa ser visto como um conjunto de partes mutuamente interagentes. Assim, esse tipo de investigação segue uma linha de pesquisa em física (física estatística), pois recentemente físicos, usando várias técnicas, têm estudado dinâmica de crescimento de diversos tipos de sistemas. Tal tipo de investigação, oriunda dos estudos de sistemas complexos, mostra, por exemplo, que o crescimento de certos sistemas é compatível com o de outros, embora possa tratar de contextos completamente distintos.

A física estatística vem avançando sobre os mais diversos campos. Sistemas que a primeira vista apresentam comportamentos aleatórios, sob um olhar cuidadoso, exibem, entretanto, certa regularidade, podendo enumerar exemplos em Geologia, Biologia, Medicina, Economia e Ciências Sociais, entre outros. Em particular, os sistemas ditos complexos são constituídos de elementos que interagem entre si e muitos desses estão diretamente relacionados com as atividades humanas, um exemplo é o sistema que nos propomos a estudar neste trabalho. Além disso, tais elementos apresentam características individuais distintas e estão organizados em grupos ou hierarquias que também podem interagir entre si. Um dos ramos da física estatística que vem se destacando nas últimas décadas é o chamado *econofísica*, em que se aplica conceitos e métodos de física estatística na análise

de mercados financeiros e outros aspectos ligados à economia [14, 15, 16]. Como exemplos específicos em que ocorrem manifestação de sistemas complexos, podemos citar o tamanho de firmas [17, 18, 19, 20], o PIB (Produto Interno Bruto) dos países [21, 22], gastos com pesquisas em universidades [23], população de pássaros [24], circulação de revistas [25] e batidas do coração [26] e caminhada humana [27].

No capítulo 1, faremos uma breve e elementar exposição de alguns tipos de leis de crescimento e distribuições de probabilidades que, de alguma forma, estão relacionados com os resultados obtidos e/ou esperados neste trabalho. Também foram dados exemplos de situações em que essas leis e distribuições se fazem presentes. De uma forma geral, esse capítulo tem como objetivo introduzir as leis de crescimento (decréscimo) exponencial e logística, assim como, as distribuições de probabilidades exponencial (Laplace), lei de potência (assintótica), gaussiana e lognormal. Concluímos o capítulo com o procedimento usado na construção de histogramas. Complementarmente ao material discutido nesse capítulo, há os apêndices A e B que dão uma fundamentação teórica às distribuições gaussiana e lognormais.

Já no capítulo 2, definiremos flutuações ou taxas de crescimento e, então, expomos alguns resultados esperados com a análise dessas flutuações, em particular o modelo de Gibrat [63] e um outro modelo estocástico proposto por Picoli Jr. [13]. Esse capítulo contém elementos básicos utilizados no estudo de séries temporais.

Dando continuidade, o capítulo 3 é direcionado a alguns dos recentes trabalhos que utilizam métodos semelhantes aos que aqui empregamos. Em particular, durante o andamento desta investigação, tomamos conhecimento de um dos trabalhos apresentados nesse capítulo. Trata-se do estudo da dinâmica de crescimento de universidades e institutos de pesquisa norte-americanos e europeus. Apesar da análise realizada se assemelhar a que pretendíamos inicialmente, mantivemos a proposta, agora com uma nova motivação: comparar a dinâmica de nossas universidades e institutos com os norte-americanos e europeus. Além do aspecto comparativo, esse capítulo servirá como apresentação da maioria das técnicas, via exemplos, que empregaremos no capítulo central desta dissertação, o capítulo 4.

Diante das etapas elaboradas até aqui, no capítulo 4 apresentaremos os resultados de nossa análise supondo, inicialmente, o mesmo tipo de crescimento para todas as instituições para que pudéssemos, de certa forma, compará-los. Em seguida, escolheremos um determinado ano para verificar a forma do histograma da distribuição de frequências e compará-la com a forma proposta pelo modelo de Gibrat, introduzido no capítulo 2. Investigaremos, também, as distribuições das taxas de crescimento e os respectivos desvios padrões. Com esses estudos, será possível caracterizar vários aspectos da dinâmica da produção científica brasileira que pode servir de base de comparação com a dinâmica de produção científica de outros países, como de outros sistemas. Como complementos a esse capítulo, adicionamos o apêndice C, em que expomos o nosso processo de coleta de dados e o apêndice D, no qual estão representados os gráficos com a produtividade anual para todas as instituições pesquisadas.

Para finalizar as idéias apresentadas até aqui, no capítulo 5 desta dissertação, apresentaremos nossas conclusões e considerações finais e de forma específica, discutiremos a tendência central de crescimento da produção científica brasileira, a restrita conexão da distribuição da produção científica com o modelo de Gibrat, a forma de tenda que tem a distribuição das taxas de crescimento e os desvios padrões das taxas de crescimento versus a produção científica para cada instituição científica do Brasil.

Por fim, para aqueles que desejam caracterizar sistemas via séries temporais de suas partes, esperamos que a forma de apresentação deste trabalho sirva como uma primeira leitura, pois nossos conceitos e técnicas foram apresentados de maneira detalhada e baseados em muitos exemplos ilustrativos.

Capítulo 1

Leis de Crescimento e Distribuições de Probabilidade

Muitas vezes para que possamos fazer o estudo de algum sistema na natureza, devemos fazer uso de aspectos determinísticos envolvendo, por exemplo, leis de crescimento, assim como fatores probabilísticos e estatísticos. De fato, notamos em diversos campos de conhecimento a presença de padrões que podem ser descritos com boa precisão por alguns tipos de distribuições de probabilidades e leis (padrões) de evolução. Exemplos das mais diversas áreas do conhecimento em que tais distribuições e leis podem ser encontradas em Física, Economia, Geologia, Ciências Sociais, Ciências Agrárias, Medicina e Biologia. Portanto, fica clara a importância do estudo dessas distribuições e leis, assim como de seu conteúdo interdisciplinar. Para ilustrar tal afirmação, trataremos de algumas distribuições de probabilidade e leis de evolução que serão utilizadas no decorrer deste trabalho, juntamente com alguns exemplos.

1.1 Crescimento Exponencial

Crescimentos (ou decrescimentos) exponenciais são caracterizados por uma “rápida” variação das grandezas envolvidas. Existe um tempo, próprio para cada sistema, em que a grandeza dobra (ou reduz a metade).

De uma maneira geral (em processos ideais), capacitores carregados quando ligados a resistores têm suas cargas elétricas decaindo exponencialmente [28]. O tempo que o capacitor leva para reduzir sua carga a metade depende das características do capacitor e do resistor.

Tomemos um outro exemplo, hipotético, em que uma população é afetada por uma epidemia e tem a taxa de mortes proporcional ao número de indivíduos, mas não há nascimento. O desenvolvimento de tal população também seguirá um padrão exponencial. Em tal caso, um decrescimento exponencial, a taxa de variação da grandeza de interesse é proporcional à própria grandeza, ou seja,

$$\frac{dz}{dt} = -\alpha z , \quad (1.1)$$

cuja solução é

$$z(t) = z_0 \exp(-\alpha t) , \quad (1.2)$$

com $z(0) = z_0$ e $\alpha > 0$ sendo constantes (veja a figura 1.1a). Outra ilustração de comportamento exponencial vem de um exemplo curioso para motivar uma aula sobre função exponencial; um professor mediu o volume de espuma de cerveja em um copo, para três diferentes tipos de cerveja. Sua conclusão foi que o volume de espuma decaiu exponencialmente [29], com um tempo próprio para cada cerveja.

Para crescimento exponencial, poderíamos imaginar também uma população de coelhos exposta a um ambiente onde não existem predadores, os alimentos são suficientes para suprir as necessidades e não há limitação de espaço. O mesmo pode ocorrer com o desenvolvimento de uma população de bactérias ou mesmo com as células de um embrião logo após a fecundação. Semelhante ao decrescimento exponencial, a taxa de variação é proporcional à grandeza estudada:

$$\frac{dz}{dt} = \alpha z . \quad (1.3)$$

Logo,

$$z(t) = z_0 \exp(\alpha t) , \quad (1.4)$$

em que, z_0 e $\alpha > 0$ são constantes (veja a figura 1.1c). Observe que agora a taxa de variação é positiva, caracterizando crescimento.

Uma propriedade do decrescimento exponencial, ilustrada na figura 1.1b é que ele se comporta como uma reta decrescente em um gráfico mono-log, ou seja, que possua as ordenadas em escala logarítmica. De fato, basta tomar o logaritmo em ambos os lados de (1.2) para obter

$$\ln z(t) = \ln z_0 - \alpha t , \quad (1.5)$$

em que $-\alpha$ corresponde à inclinação da reta. O mesmo ocorre com o crescimento exponencial, veja a figura 1.1d, entretanto a reta no gráfico mono-log será crescente:

$$\ln z(t) = \ln z_0 + \alpha t . \quad (1.6)$$

1.1.1 Exemplos

Com o intuito de ilustrar a ampla diversidade de assuntos em que estão presentes os crescimentos exponenciais, além dos já citados, vamos elencar alguns outros exemplos específicos que estão ligados às atividades humanas.

São exemplos de sistemas que se comportam exponencialmente, o aumento do número de sequências de DNA contidas no *GenBank*, um banco de dados do *National Center for Biotechnology Information (NCBI)* [30], o número de pacotes do sistema operacional *FreeBSD* [32], o número de idiomas em que a bíblia foi publicada [33], a *Lei de Moore* sobre o aumento da capacidade de processamento dos computadores [35], entre outros.

GenBank

O avanço da biotecnologia tem permitido recentemente o seqüenciamento de cadeias de DNA de diversas espécies. Visto a enorme diversidade dos seres vivos existentes, fez-se necessário organizar os resultados em algum banco de dados que permitisse à comunidade científica um acesso fácil, rápido e inteligente. Um destes bancos de dados foi construído pelo Instituto Nacional de Saúde dos EUA (*National Institute of Health*), o NCBI [31]. A partir daí houve a necessidade da socialização dos dados.

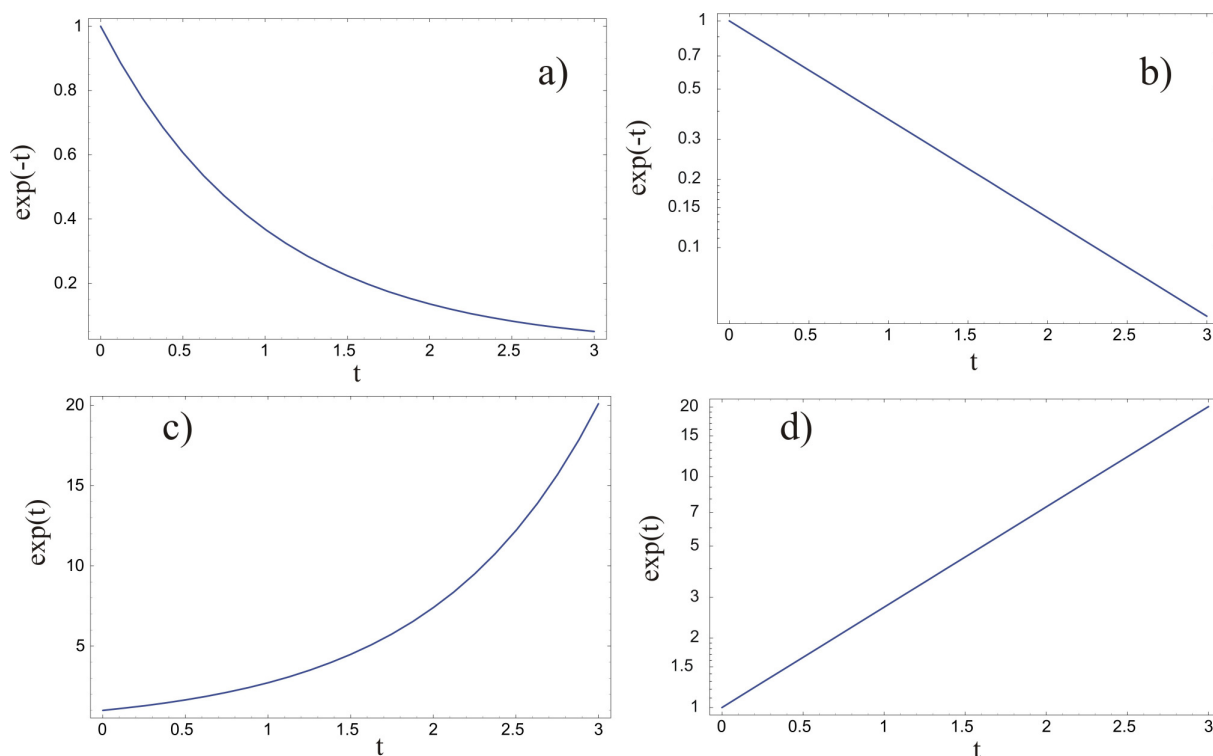


Figura 1.1: Em a), vemos a representação do decrescimento exponencial em escala linear e, em b), o mesmo decrescimento em um gráfico mono-log. Em c), o crescimento exponencial com escala linear é exibido, e em d), com escala mono-log.

Foi criada uma força tarefa internacional para montar um banco de dados de seqüências de DNA, a INSDC (*International Nucleotide Sequence Database Colaboration*). Desse banco de dados fazem parte o NCBI (*National Center for Biotechnology Information*), o EMBL (*European Molecular Biology Laboratory*) e o DDJB (*DNA Data Bank of Japan*). Cada um desses centros possibilita a submissão individual de seqüências de DNA e trocam informações entre si diariamente, sendo que todos os três possuem informações atualizadas de todas as seqüências disponíveis para os pesquisadores.

O banco de dados do NCBI, o GenBank, apresentou um crescimento exponencial do número seqüências armazenadas ao longo desses anos, veja figura 1.2. Em agosto de 2005, o INSDC completou um total de 100 milhões de bases.

FreeBSD

BSD é a sigla para *Berkeley Software Distribution*. Trata-se de uma criação feita por estudantes na Universidade da Berkeley, da Califórnia, para aprimorar e acrescentar algumas funcionalidades ao sistema operacional Unix. A idéia inicial foi sofrendo mudanças rapidamente. Tantas que chegou a um ponto em que o BSD virou, praticamente, um novo sistema operacional.

O BSD possui uma licença própria. Por meio dessa licença, qualquer pessoa/empresa pode fazer o que quiser do código, inclusive, alterá-lo para a criação de produtos fechados. Os três sistemas operacionais BSD's mais conhecidos e que seguem esse padrão de licença são: OpenBSD, FreeBDS e NetBSD.

Visto que qualquer pessoa pode alterar e/ou criar novos produtos para esses sistemas

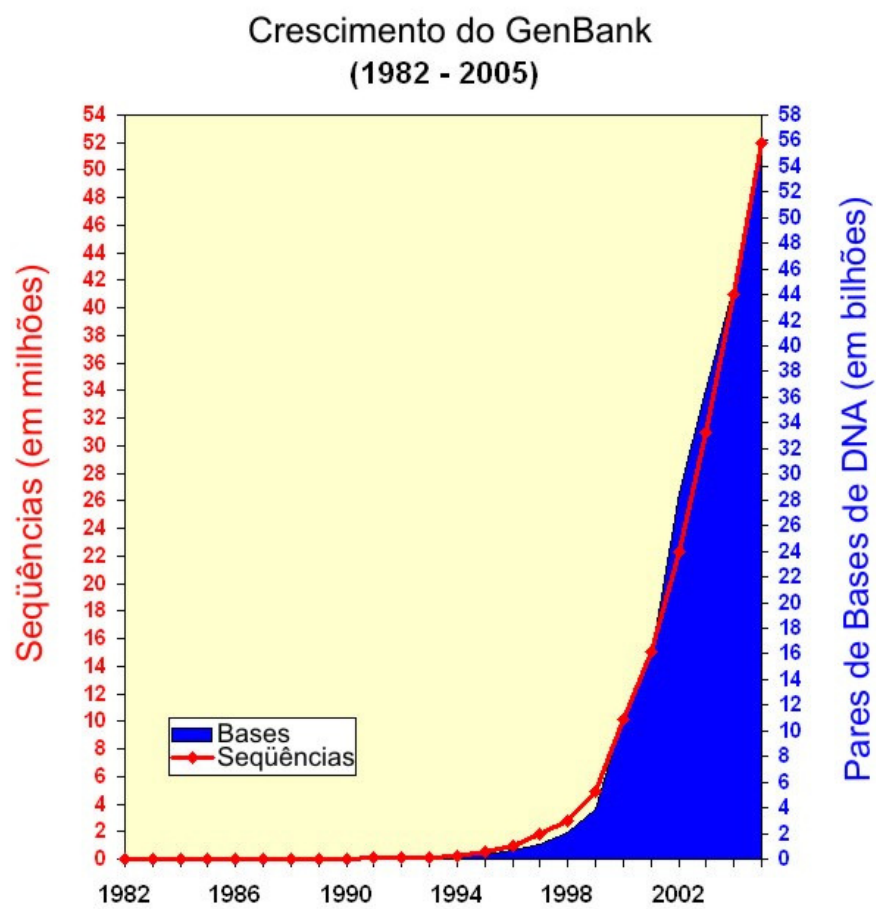


Figura 1.2: Crescimento do Banco de Dados GenBank. Na escala à esquerda, é apresentado o número de seqüências catalogadas (em milhões), e à direita o número de pares de bases (em bilhões). Imagem retirada do *site* do NCBI [31].

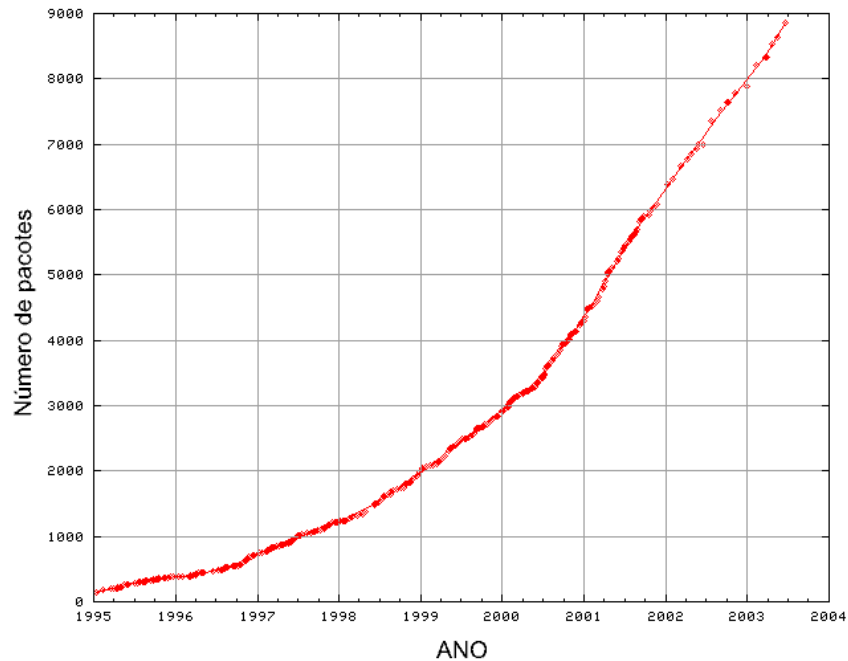


Figura 1.3: Crescimento do número de pacotes do sistema operacional FreeBSD. Figura retirada da ref. [32].

operacionais, foi se acumulando ao longo do tempo uma quantidade notável de pacotes¹ do FreeBSD. Esse crescimento mostrou ser exponencial ao longo destes anos [32], veja a figura 1.3.

Bíblia

A bíblia foi escrita por cerca de 40 homens, em três idiomas, o *grego*, o *hebraico* e o *aramaico*, durante cerca de 1600 anos. Ela foi, inicialmente, dividida em duas grandes partes, o *Antigo Testamento* e o *Novo Testamento*, cada uma com vários livros. E, mais tarde, subdividida ainda em capítulos e versículos².

Os primeiros exemplares da bíblia produzida eram manuscritos e levavam vários meses para ficarem prontos. Por isso, seu preço era bastante elevado e poucos tinham acesso. A bíblia só passou a ser produzida em larga escala com a invenção da prensa gráfica por *Johannes Gutenberg*. Com o avanço da tecnologia, ficou enorme a praticidade e vários são os idiomas em que fora publicada. A figura 1.4 representa a evolução exponencial do número de idiomas em que a bíblia foi publicada [33].

Lei de Moore

Na natureza, os crescimentos exponenciais de longa duração são raros. Por exemplo, quando se trata de dinâmica das populações, os padrões exponenciais, quando ocorrem,

¹Entende-se por pacote o “conjunto de programas e manuais de computador que abrangem todos os aspectos de uma tarefa especial, como folha de pagamento, controle de caixa, etc” (*Michaelis eletrônico*).

²Acredita-se que a divisão em capítulos tenha surgido no século XIII e a divisão em versículos no século XVI.

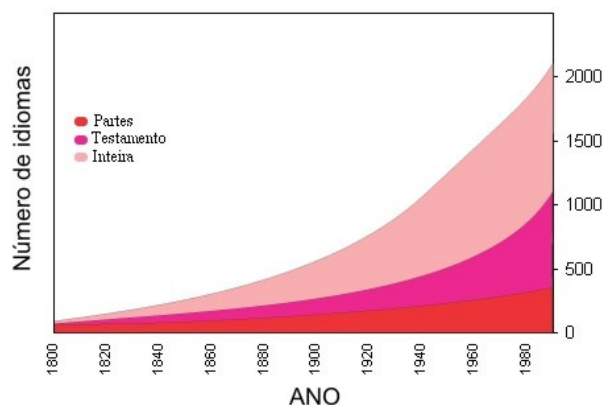


Figura 1.4: Crescimento dos idiomas em que a bíblia foi traduzida e publicada, em apenas alguns livros, um testamento ou inteira. Figura retirada da ref. [33].

são de curta duração, visto que os recursos são limitados. Há então, uma diminuição do ritmo de crescimento, podendo até chegar a uma estagnação [34].

O mesmo não parece ocorrer com informática e tecnologia. Elas vêm mantendo um ritmo de crescimento exponencial por vários anos. *Gordon Moore* observou que os computadores tinham um desenvolvimento muito rápido e dobravam a capacidade de processamento a cada dois anos, enquanto o custo permanece constante [35]. Os dados obtidos com o desenvolvimento de mais e mais processadores concordam com a proposta de *Moore* por mais de 30 anos. Veja a figura 1.5.

1.2 Crescimento Logístico

Uma população de bactérias que se multiplica, por exemplo, exibe inicialmente um rápido crescimento, mas tende à estagnação, pois, logo elas começam a competir entre si por espaço e alimento. Nesse tipo de situação, ocorrem crescimentos que inicialmente são exponenciais, mas após um certo período deixam de ser. Neste contexto, a lei de crescimento (distribuição) logística tem um papel especial.

Numa situação mais geral que aquela dada pela equação (1.3), tem-se $dz/dt = f(z)$. Se para $f(z)$ for considerada sua série de Taylor, porém até segunda ordem, $f(z) \simeq \alpha z - \beta z^2$, a equação diferencial resultante é

$$\frac{dz}{dt} = \alpha z - \beta z^2, \quad (1.7)$$

em que os dois fatores α e β são constantes positivas.

Note que $\alpha > 0$ garante o crescimento exponencial para z pequeno e $\beta > 0$ conduz a um valor finito z quando t for grande. Esses aspectos podem ser verificados diretamente da solução da equação (1.7):

$$z(t) = \frac{\alpha}{\beta + C e^{-\alpha t}}, \quad (1.8)$$

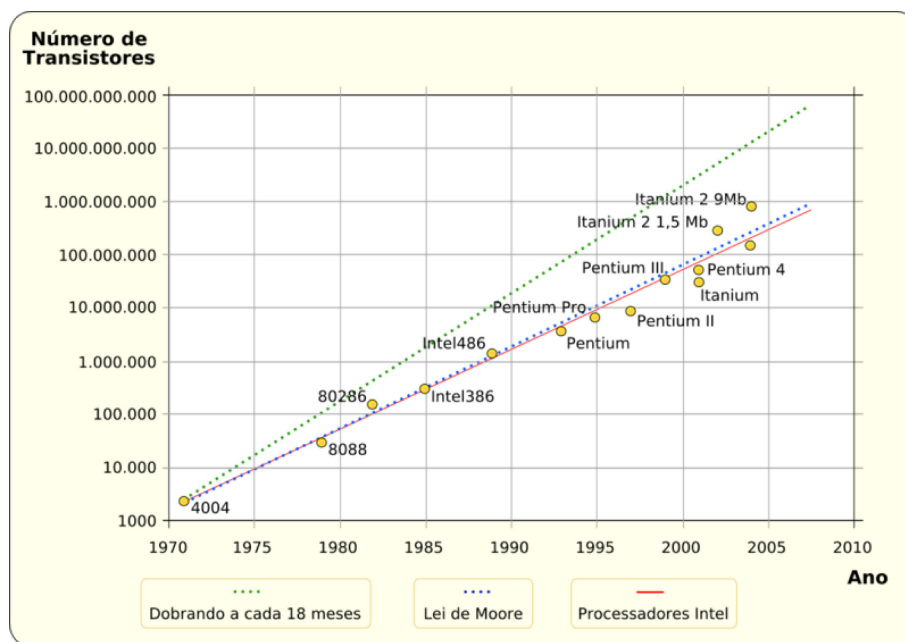


Figura 1.5: Crescimento da capacidade de processamento, associada ao número de transistores. Note que o gráfico é mono-log e por isso uma reta.

com $z(0) = z_0$ e $C = \alpha/z_0 - \beta$. Realmente, se $\beta \ll Ce^{-\alpha t}$, então

$$z(t) = (\alpha/C)e^{\alpha t}, \quad (1.9)$$

assim chegamos ao comportamento exponencial, como ilustrado na figura 1.6. Para $\alpha t \gg 1$, obtemos $z(t) = \alpha/\beta$, e isso ilustra o momento em que não há mais crescimento. A equação (1.7) é usualmente referida como equação logística.

A forma do gráfico de uma distribuição logística é representada na figura 1.6. Nota-se a tendência inicial de crescimento rápido (exponencial) seguido de uma diminuição deste ritmo. Para esse tipo de distribuição não é comum a utilização de escalas logarítmicas.

1.2.1 Exemplos

Com boa precisão, distribuições logísticas são encontradas em várias situações, como na estimativa de crescimento da população mundial [36], no estudo da resposta de insetos ao uso de inseticidas [37], número de infectados vivos com HIV [38], gastos do governo [39] e crescimento de frutos de manga [40, 41, 42].

População Mundial

Malthus defendia a idéia de que haveria um grande problema mundial de falta de alimentos [43]. Ele acreditava que a população cresceria exponencialmente (geometricamente), enquanto que a produção de alimentos não conseguiria manter o mesmo crescimento, tendo um crescimento linear. Anos mais tarde, *Verhulst* propôs uma alteração nas idéas de *Malthus*, levando em conta essa possível competição por alimentos [34]. Essa nova idéia foi defendida bravamente, anos mais tarde, por *Pearl* [44]. Foi então que surgiu o modelo logístico para explicar a dinâmica das populações, sugerindo que seu crescimento segue uma curva logística.

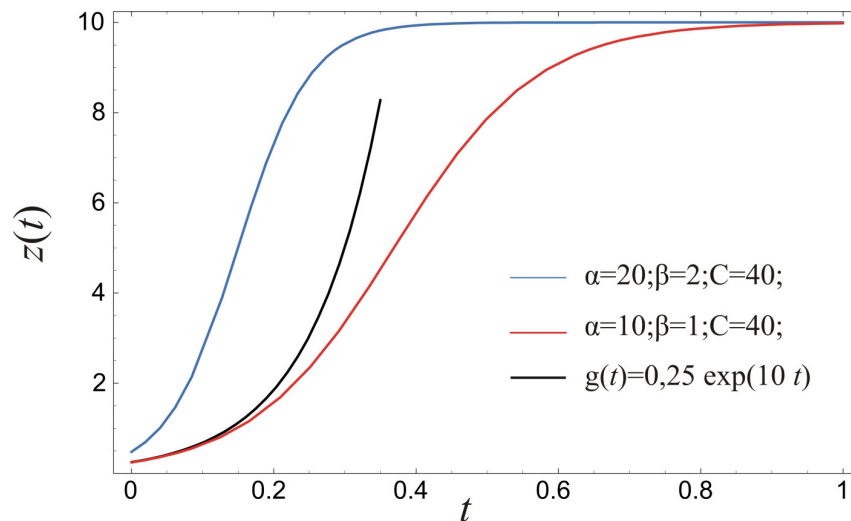


Figura 1.6: As curvas logísticas são representadas pela equação (1.8). Em preto, temos uma curva exponencial para ilustrar a tendência inicial do crescimento logístico, note que ela satisfaz a equação (1.9).

A figura 1.7 representa uma estimativa para a população mundial ao longo de 100 anos (1950-2050), segundo o *US Census Bureau*[36], dos Estados Unidos da América.

Insetos

Simplificadamente, a resposta dos insetos aos inseticidas é bastante simples: eles morrem ou não morrem. Em um estudo de um tipo específico de inseticida, *rotenone*, ele foi aplicado em uma espécie de pulgão, *Macrosiphoniella sanborni*. O interesse do pesquisador estava em determinar a relação entre as doses aplicadas e a proporção de insetos mortos.

O resultado encontrado mostrou que essa relação é bem representada por um comportamento logístico [37]. Veja a figura 1.8.

Aids

O *Vírus da Imunodeficiência Humana*, conhecido como HIV (sigla originada do inglês: *Human Immunodeficiency Virus*), pertence à classe dos retrovírus e é causador da AIDS (*Síndrome da Imunodeficiência Adquirida*).

Ter o HIV não é a mesma coisa que ter a AIDS. Significa que, no sangue, foram detectados anticorpos contra o vírus. Há muitas pessoas soropositivas que vivem durante anos sem desenvolver a doença, no entanto, podem transmitir aos outros o vírus que trazem consigo.

As células do sistema imunológico de uma pessoa infectada pelo vírus, que desenvolveu a síndrome, começam a funcionar com menos eficiência e, com o tempo, a habilidade do organismo em combater doenças comuns diminui, deixando a pessoa sujeita ao aparecimento de vários outros tipos de doenças e infecções [45].

Embora o número de novos casos tenha diminuído bastante, em geral, o número de portadores de AIDS ainda cresce. Como exemplo, tomemos o caso da África sub-Saara, em que percebemos um crescimento segundo uma curva logística [38]. Veja a figura 1.9.

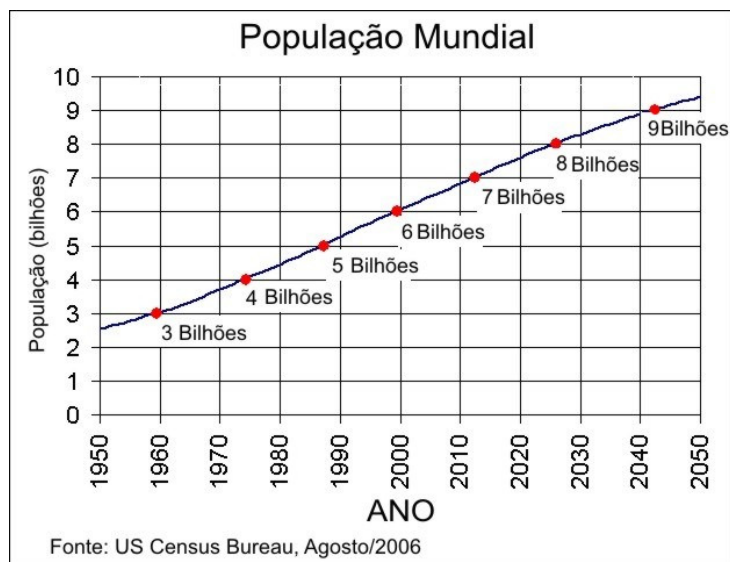


Figura 1.7: Estimativa do *US Census Bureau* para a população mundial, em bilhões, ao longo dos anos.

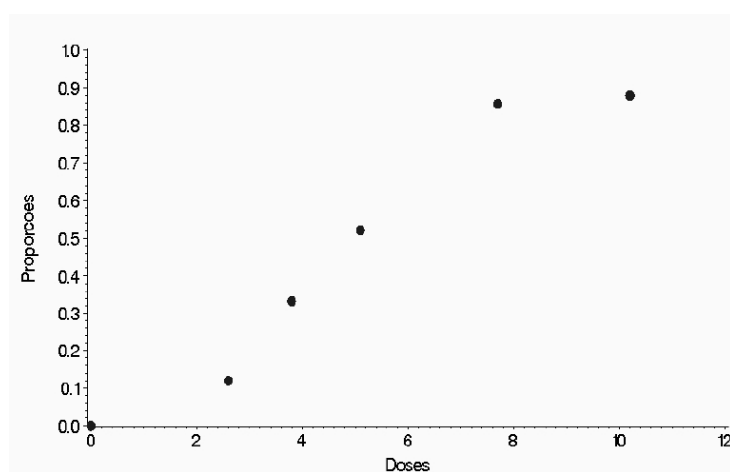


Figura 1.8: Relação entre proporção de insetos mortos e a dose aplicada.

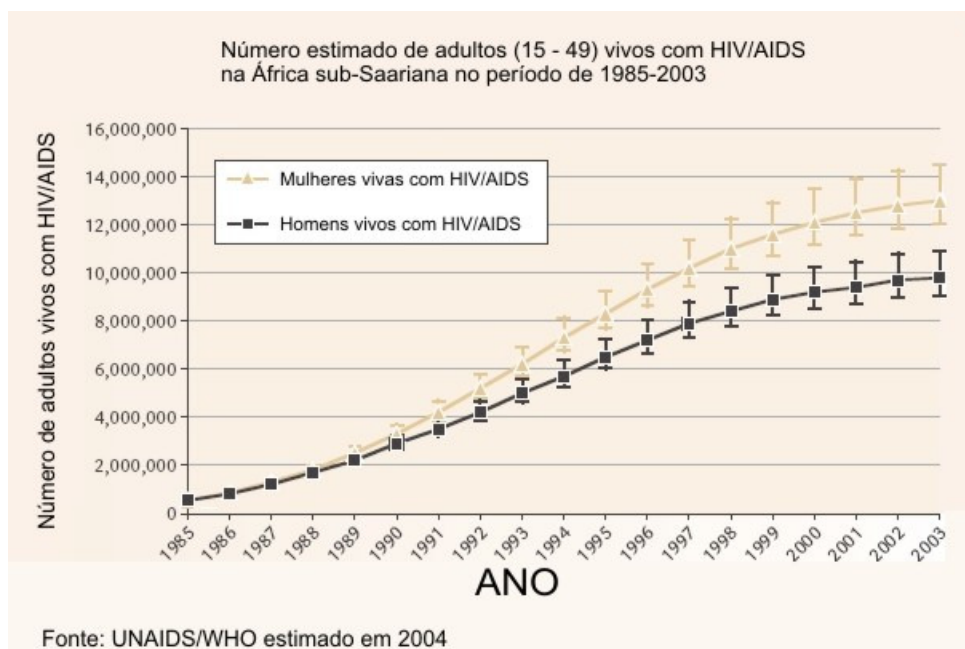


Figura 1.9: Estimativa do número de casos de AIDS na África sub-Saariana, para adultos, homens e mulheres, com idade entre 15 e 49 anos. Os quadrados representam os homens e os triângulos as mulheres. Fonte UNAIDS[38].

Gastos do Governo

Considera-se os gastos do governo, para um estudo particular, como sendo gastos em justiça e ordem pública, infraestrutura, sistema militar, sistema de educação, sistema de saúde, programas sociais, suporte a empresas, políticas especiais (anti-drogas, estrangeiros, anti-terrorismo etc), entre outras.

Foi realizado um estudo em alguns países, em particular, Estados Unidos da América, França, Alemanha, Itália e Inglaterra, sobre os gastos do governo. A conclusão foi que a razão entre os gastos e a arrecadação segue a forma de uma curva logística [39]. Tal estudo se estendeu por mais de 100 anos. Para estes países, no final do século XIX, essa razão situava-se entre 5% e 10%. Já no final do século XX, achava-se entre 35% e 65%. Veja a figura 1.10.

Frutos

Ao estudar a evolução dos frutos de manga *cv. Haden*, amostras foram colhidas em datas correspondendo aos 15, 23, 30, 37, 42, 47, 55 e 73 dias depois a floração. Após a colheita, os frutos amostrados foram levados para laboratório e imediatamente pesados individualmente para a obtenção da *massa fresca*. Para a obtenção da *massa seca*, os frutos foram seccionados em pequenos pedaços e colocados para secar em uma estufa de ventilação forçada com temperatura de $65\text{ }^{\circ}\text{C} \pm 5\text{ }^{\circ}\text{C}$. Os frutos foram mantidos em estufa até atingirem peso constante [40]. A figura 1.11 mostra como foi a evolução da massa seca nos respectivos dias de colheita. Estudos semelhantes já haviam sido feitos [41, 42] e as conclusões foram as mesmas.

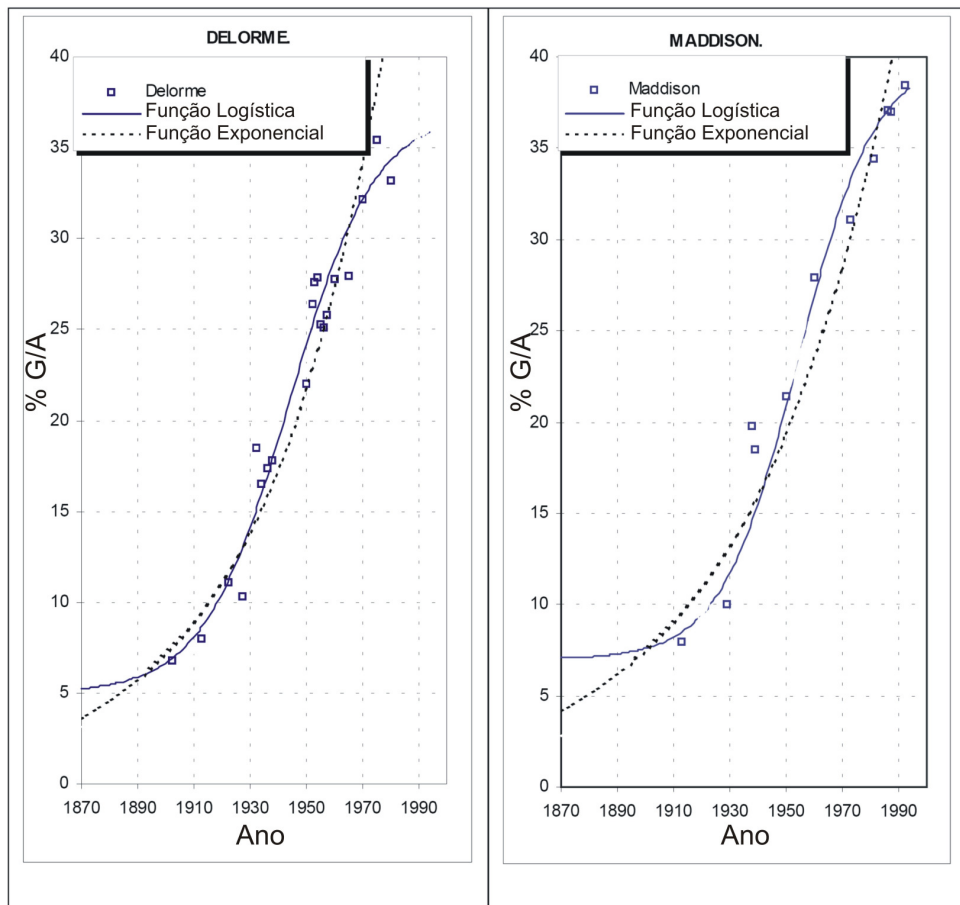


Figura 1.10: Figura retirada e adaptada da referência [39]. Aqui, representa-se a razão entre os gastos (G) e a arrecadação (A) do governo norte-americano ao longo dos anos. Esse crescimento segue uma curva logística.

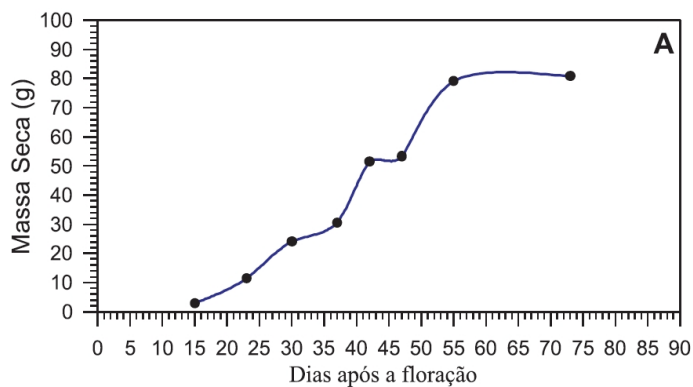


Figura 1.11: Figura retirada da referência [40]. A evolução da massa seca do fruto da manga *cv. Haden* segue uma curva logística.

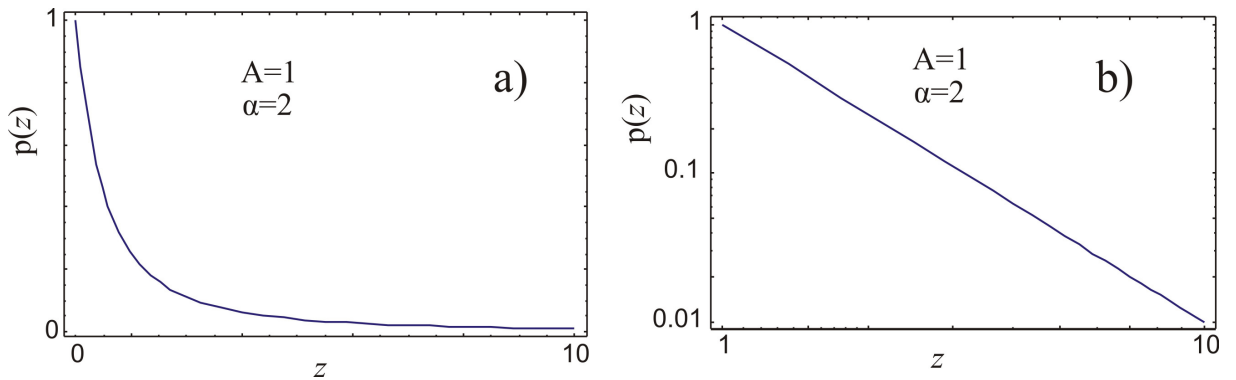


Figura 1.12: Em a), mostramos a representação da equação (1.10) em escala linear. Os parâmetros estão indicados na figura. Em b), a escala logarítma é utilizada em ambos os eixos.

1.3 Distribuição do Tipo Lei de Potência

É comum encontrarmos, no estudo de “sistemas complexos”, variáveis que se distribuem de acordo com uma lei de potência,

$$p(z) = Az^{-\alpha} , \tag{1.10}$$

que apresenta um decaimento mais suave que um exponencial. Particularmente, em economia, essa distribuição é usualmente conhecida como distribuição de *Pareto* [46].

As distribuições do tipo leis de potência também são chamadas de distribuições livres de escala. Isso porque a lei de potência é uma distribuição que apresenta a mesma forma para qualquer escala utilizada. Uma distribuição livre de escala deve obedecer a relação

$$p(bz) = g(b)p(z) , \tag{1.11}$$

para um b qualquer. Note que a lei de potência (1.10) satisfaz essa condição, pois a equação (1.10) conduz a

$$\begin{aligned} p(bz) &= A(bz)^{-\alpha} \\ &= b^{-\alpha}p(z) , \end{aligned} \tag{1.12}$$

com $g(b) = b^{-\alpha}$.

A propriedade de que as leis de potência se comportam como retas em gráficos log-log (escala logarítmica em ambos os eixos) é bastante conveniente quando analisamos dados experimentais, pois a comparação de dados com uma reta tem um apelo visual direto. Para obter a relação entre A e α com os coeficientes da reta, basta tomar o logaritmo em ambos os lados de (1.10):

$$\ln p(z) = \ln A - \alpha \ln z , \tag{1.13}$$

em que $\ln A$ é o coeficiente linear e $-\alpha$ corresponde a inclinação da reta. A figura 1.12 representa tais distribuições em escala linear e logarítmica (log-log).

1.3.1 Exemplos

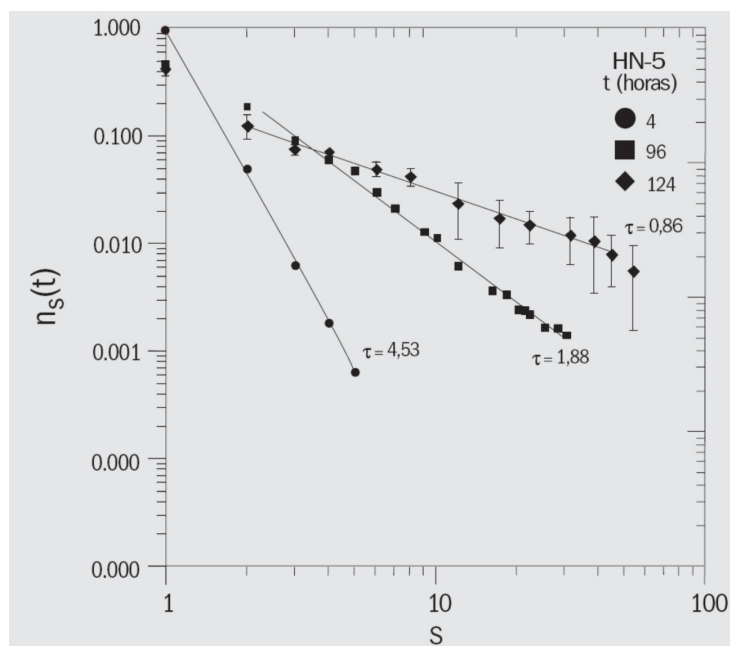


Figura 1.13: Figura retirada da referência [47]. A distribuição de tamanhos de agregados, n_s , segue uma função do tipo lei de potência com o aumento do número de células, S , representada por retas em gráficos log-log. Aqui $-\tau$ é o expoente da lei de potência.

Assim como nos casos anteriores, apresentaremos alguns exemplos ilustrativos ressaltando a diversidade de situações em que ocorrem leis de potência. Tais distribuições são encontradas, entre outros exemplos, para o tamanho de agregados de alguns tipos de células [47], conectividade entre grandes redes através da *internet* [48], número de blecautes ocorridos nos sistemas de transmissão de energia elétrica norte-americano [49] e no número de ataques terroristas [50].

Agregados de Células

Foi realizado um estudo sobre o padrão de crescimento de alguns tipos de células. Em especial, as células do tipo HN-5, que são células cancerosas, durante sua multiplicação, costumam se agrupar formando agregados. O tamanho desses agregados apresentam um desenvolvimento regido por uma lei de potência, como ilustra a figura 1.13. Note que o gráfico é do tipo log-log (escala logarítmica em ambos os eixos). A função de distribuição de tamanhos de agregados, $n_s(t)$, foi definida como sendo a fração de agregados celulares consistindo de s células em um tempo t , veja a referência [47].

Redes

Redes são caracterizadas por elementos interligados entre si. Em se tratando de informática, um computador pode se ligar diretamente a outros, ou ainda, todos os computadores de uma instituição podem ser ligados a um outro, o servidor. Para acessar um *site* qualquer na *internet*, muitos destes servidores também se conectam.

É interessante notar que, independente do sistema e da identidade de seus constituintes, a probabilidade $p(k)$ que cada vértice da rede interaja com outros k vértices decaem como uma lei de potência [48], seguindo

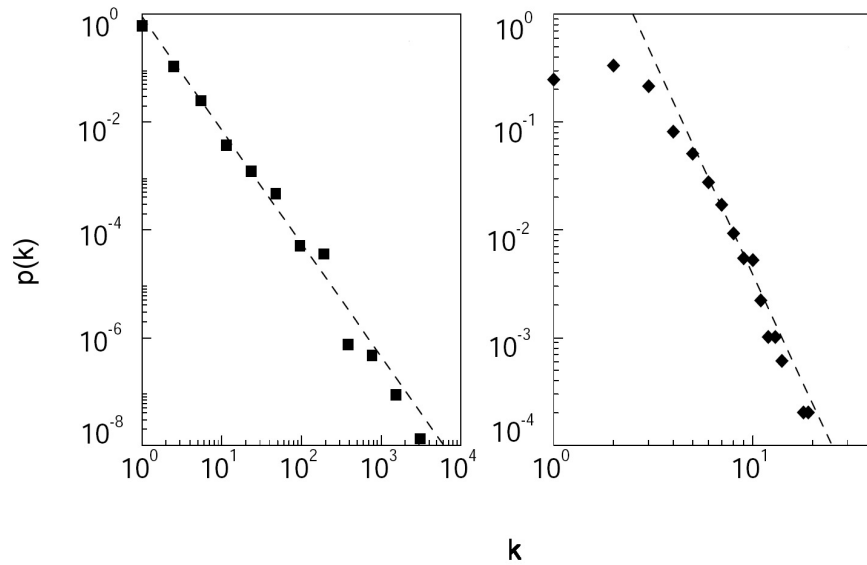


Figura 1.14: Figura retirada e adaptada da referência [48]. Representa-se a probabilidade, $p(k)$, de cada vértice interagir diretamente com outros k vértices. Em a), o sistema observado é a *internet*, e a linha pontilhada é uma função $p(k) \propto k^{-2,1}$. Já em b), representa-se a rede das linhas de distribuição de energia elétrica e a linha pontilhada é uma função $p(k) \propto k^{-4}$.

$$p(k) \propto k^{-\gamma} . \quad (1.14)$$

O valor de γ é característico de cada sistema, no caso da *internet* $\gamma_{www} = 2,1$, já para as linhas de transmissão de eletricidade $\gamma_{linhas} = 4$. A figura 1.14 trata destes exemplos.

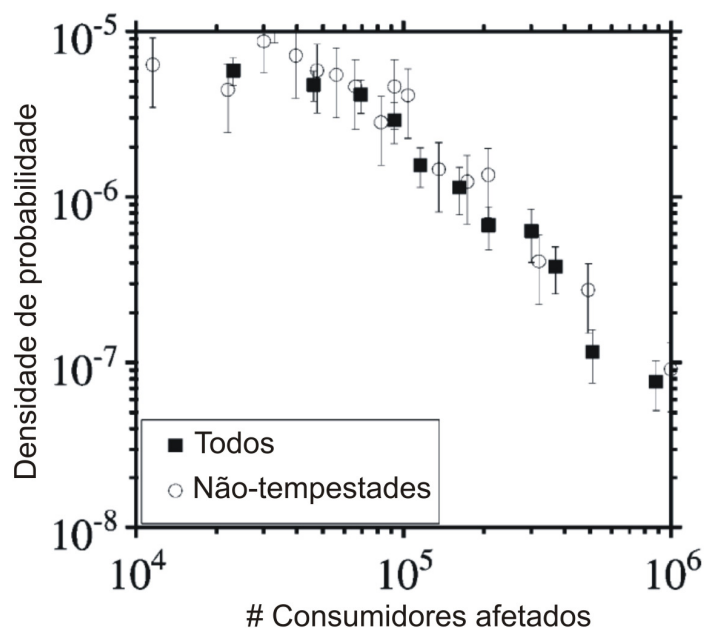


Figura 1.15: Figura retirada e adaptada da referência [49]. Distribuição de probabilidades para os blecautes ocorridos nos Estados Unidos da América.

Blecautes

As redes de transmissão de eletricidade norte-americana são sistemas um tanto complexos e comumente trabalham próximas aos limites operacionais. Distúrbios ou blecautes nesses sistemas de transmissão conduzem a sérias conseqüências. Os blecautes podem ser atribuídos a causas específicas, tais como raios, tempestades de neve, falha dos equipamentos e queda de árvores.

Uma análise da série temporal dos blecautes ocorridos mostra que há uma relação do tipo lei de potência entre a probabilidade de ocorrer um blecaute e o número de linhas afetadas (tamanho do blecaute) [49]. A figura 1.15 representa tal afirmação.

Ataques Terroristas

Ultimamente, muito se ouve notícias sobre ataques terroristas. O Oriente Médio sofre diariamente com homens-bomba, carros-bomba e até um “hilário” caso de bicicleta-bomba. Casos bem menos freqüentes nos Estados Unidos da América, Espanha, Inglaterra, entre outros países mais desenvolvidos, também ocorrem.

Um estudo estatístico do número de vítimas (mortos ou feridos) também conduz a uma lei de potência, representada pela freqüência e “dureza” em que ocorrem os ataques [50]. A figura 1.16 ilustra este tipo de estudo, comparando diversos tipos de ataques.

1.4 Distribuição Gaussiana

Alguns sistemas apresentam grandezas que se distribuem simetricamente em torno de um pico central. Freqüentemente, histogramas (vide sessão 1.6) de tais grandezas podem ser representados por uma distribuição gaussiana. É uma ferramenta útil para os

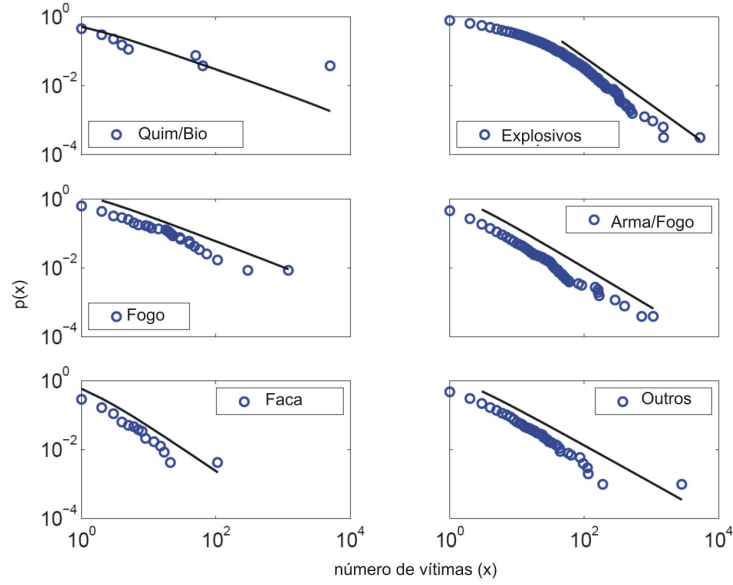


Figura 1.16: Figura retirada e adaptada da referência [50]. A distribuição acumulada (veja a discussão correspondente a equação (3.7)) do número de vítimas para diversos tipos de ataques.

estatísticos, por exemplo, ao analisar um estudo sobre o preço de um determinado produto. Imagine que o estudo consistisse em determinar o preço de uma garrafa pequena de um determinado refrigerante. Suponha que fossem encontrados inúmeros preços diferentes, mas o preço médio foi de R\$ 1,10. É bem provável que encontremos esse refrigerante com preço de R\$ 1,20, pouco provável que encontremos preços de R\$ 2,00, e que talvez nunca encontrássemos um preço de R\$ 5,00.

Uma distribuição gaussiana é dada por

$$p(z) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(z - z_0)^2}{2\sigma^2}\right], \quad (1.15)$$

em que $\langle z \rangle = z_0$ é o valor médio e $\sigma = \sqrt{\langle (z - \langle z \rangle)^2 \rangle}$ corresponde ao desvio padrão. A conexão entre distribuição gaussiana e a soma de variáveis aleatórias independentes está no apêndice A. As distribuições gaussianas ocorrem como consequência da soma de muitos eventos aleatórios independentes (Teorema Central do Limite). E justamente por esse motivo que ela surge em muitos contextos na ciência em geral, sendo uma grande tendência na natureza.

A distribuição gaussiana, também conhecida como distribuição normal, apresenta o gráfico com a forma de sino. O pico central corresponde ao valor médio da distribuição e a largura está associada ao desvio padrão. A figura 1.17a mostra esse gráfico.

Uma propriedade observada nas distribuições gaussianas é que, em gráficos mono-log, elas exibem a forma de uma parábola (veja a figura 1.17b). Basta aplicar o logaritmo em ambos os lados da equação (1.15) e assim constatamos que

$$\ln p(z) = \ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \frac{(z - z_0)^2}{2\sigma^2}. \quad (1.16)$$

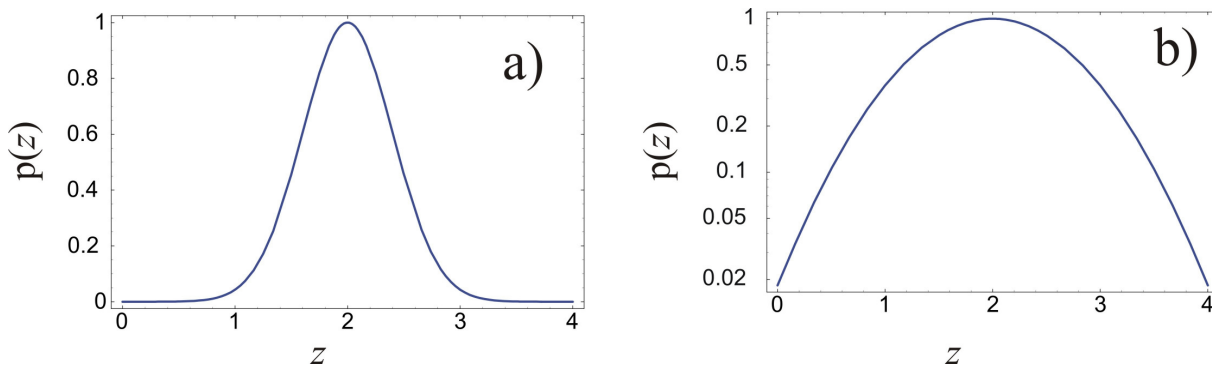


Figura 1.17: O gráfico, em escala linear, de uma distribuição gaussiana tem a forma de um sino, como vemos em a). Quando utilizamos escala logaritma no eixo das ordenadas obtemos uma parábola, como visto em b).

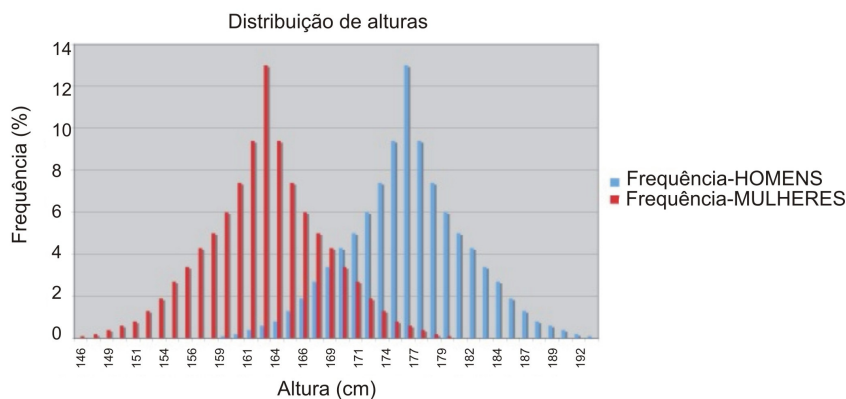


Figura 1.18: Histograma representando a distribuição de frequências da altura de seres humanos. Note que o valor médio é diferente para homens e mulheres.

1.4.1 Exemplos

Os exemplos são novamente inúmeros, distribuição de velocidades das moléculas de um gás [51], a altura dos humanos [52], os erros em medidas experimentais, na distribuição do número de crianças com deficiência de ferro e anemia na cidade de Vitória - ES [53], entre outros.

Altura de Humanos

Nota-se que ao medir a altura de um conjunto de pessoas os resultados encontram-se dispostos ao redor de um pico central. Essa foi a conclusão de *Joiner* ao estudar a distribuição da altura das pessoas [52]. O padrão, distribuição gaussiana, é o mesmo para homens e mulheres, embora o valor médio seja diferente. A figura 1.18 representa um histograma das medidas de alturas em seres humanos.

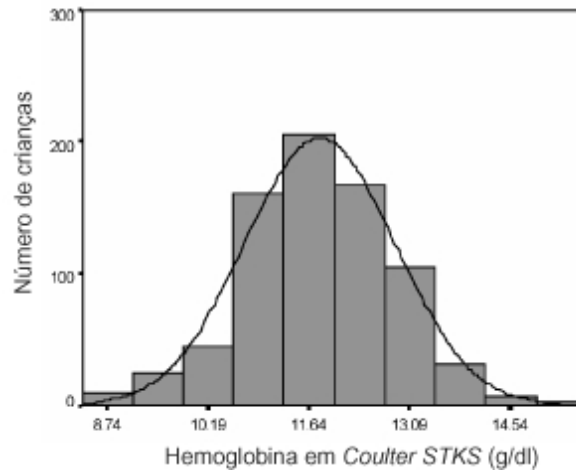


Figura 1.19: Figura retirada e adaptada da referência [53]. A distribuição de frequência do número de crianças com a dada concentração de hemoglobina no sangue.

Deficiência em ferro e anemia

Houve um estudo realizado com 760 crianças, com idade de 6 meses a 7 anos, selecionadas aleatoriamente nos Centros Municipais de Educação Infantil na cidade de Vitória - ES. Foi então medida a taxa de hemoglobina presente no sangue. O resultado seguiu uma distribuição gaussiana. Veja a figura 1.19.

1.5 Distribuição Lognormal

Em estatística, as distribuições lognormais são distribuições de probabilidades de variáveis aleatórias na qual seus logaritmos têm distribuição normal (gaussiana). Portanto, se z obedece uma distribuição lognormal, então $\ln(z)$ é normalmente distribuído. Alternativamente, se z é uma variável aleatória com distribuição normal, $\exp(z)$ segue uma distribuição lognormal.

A distribuição lognormal é dada por

$$p(z) = \frac{1}{z\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln z - \langle \ln z \rangle)^2}{2\sigma^2}\right] \quad (1.17)$$

e o detalhamento da conexão desta distribuição com a gaussiana está apresentado no apêndice B.

Na figura 1.20, gráficos para distribuições lognormais são apresentados. Notamos em tal distribuição uma calda mais alongada se comparada com a distribuição gaussiana.

1.5.1 Exemplos

Novamente os exemplos são muitos. Podemos citar trabalhos em Geologia em que foi descoberto que na crosta terrestre a concentração de elementos e seus agentes radioativos geralmente seguem a distribuição lognormal [54]. Ou ainda em Medicina, descobriu-se que os períodos de latência de doenças infecciosas se ajustam a distribuições lognormal [55, 56, 57, 58]. Em Ecologia, constatou-se que no reino animal (pássaros, peixes, insetos)

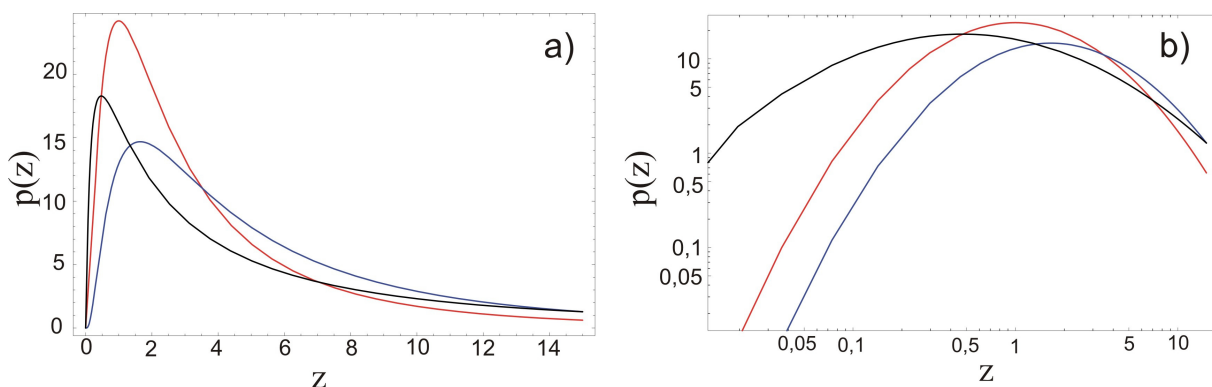


Figura 1.20: A distribuição lognormal em gráfico com escala linear em a) e em log-log em b), onde o gráfico tem a forma de uma parábola. Em vermelho, mostra-se a distribuição com $\mu = 1,0$ e $\sigma = 1,0$, já em azul, $\mu = 1,5$ e $\sigma = 1,0$ e finalmente em preto, $\mu = 1,5$ e $\sigma = 1,5$.

e vegetal, muitas de suas comunidades seguem a distribuição lognormal [59] para o número de indivíduos por espécie. Em Ciências Sociais, constatou-se que seguem essa distribuição a idade de casamento, tamanho de fazendas e rendimentos [60]. Já em Economia, há um exemplo de grande popularidade em física estatística que se refere ao crescimento de firmas nas quais seus tamanhos apresentam, com boa aproximação, distribuição lognormal [63]. Para concluir a nossa lista de exemplos, ilustramos graficamente a indicação de uma lognormal para a distribuição dos tempos de bola fora de jogo (parada) em futebol [61] (veja a figura 1.21).

1.6 Histogramas

A representação das distribuições geralmente é feita por meio de histogramas. Eles consistem em representações gráficas da distribuição de freqüências de um conjunto de dados, sendo ferramentas úteis para a apresentação dos mesmos em uma forma em que se torna mais evidente a tendência central e a dispersão (desvio padrão). De forma simplificada, podemos dizer que distribuições de freqüências tratam-se de funções em que, como variável independente, nas abscissas, temos valores que os dados assumem e como variável dependente, nas ordenadas, a freqüência com que tal valor se repete.

O procedimento mais comum para fazer um histograma é considerar a variável independente z subdividida em intervalos (janelas) Δz_i 's. A seguir, contam-se os eventos que correspondem a cada intervalo Δz_i , denotando-os por N_i , respectivamente. Então, se constrói o histograma: ao eixo vertical relaciona-se cada N_i e ao eixo horizontal os respectivos Δz_i 's.

Se o conjunto de dados vier de uma distribuição de probabilidades $p(z)$, podemos relacioná-la com a freqüência N_i e o intervalo Δz_i via a relação

$$p(z_i)\Delta z_i \simeq \frac{N_i}{N}. \quad (1.18)$$

Tal relação se justifica, pois a freqüência relativa N_i/N (com $N = \sum_i N_i$) tende para a probabilidade $p(z_i)\Delta z_i$ a medida que os N_i 's são suficientemente grandes e os Δz_i 's pequenos.

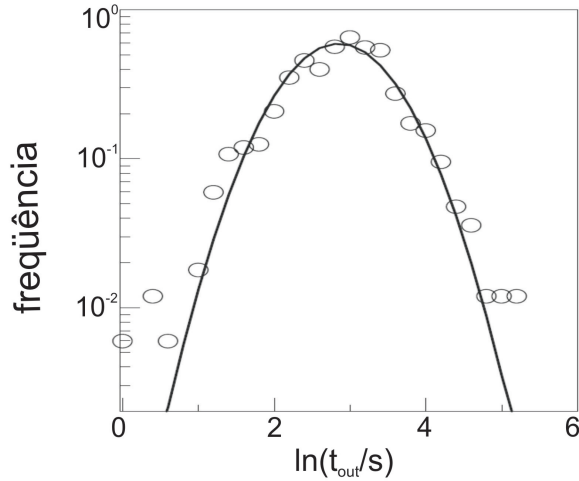


Figura 1.21: Figura retirada e adaptada da referência [61]. A distribuição lognormal em futebol, em que $\ln(t_{out}/s)$ representa o logaritmo do tempo que fica parado um jogo de futebol devido a faltas, arremessos laterais, etc. A curva contínua representa uma parábola e os círculos são dados observacionais.

Um exemplo típico de uso da relação 1.18,

$$p(z_i) \simeq \frac{N_i}{N\Delta z_i}, \quad (1.19)$$

está apresentado na figura 1.22. Nesse caso, consideramos a distribuição de Laplace (também referida neste trabalho como exponencial),

$$p(z) = \frac{1}{\sqrt{2}\sigma} \exp\left(-\frac{\sqrt{2}|x - \mu|}{\sigma}\right), \quad (1.20)$$

em que σ é o desvio padrão e μ a média, que na figura 1.22 tem $\sigma = 1$ e $\mu = 0$.

Para obter a figura 1.22, usamos um gerador de números aleatórios que obedecem a distribuição de Laplace, com desvio padrão unitário. Geramos 10^4 números aleatórios e usamos todos Δz_i 's iguais, isto é, $\Delta z_i = \Delta z$ para todo i , com $\Delta z = 0,34$.

Neste texto, é comum dizer que a distribuição de Laplace (exponencial do módulo) tem forma de tenda (*tent shape*) por razão óbvia ao olharmos as figuras 1.22b e 1.22c.

Percebemos ainda que as figuras 1.22b e 1.22c apresentam flutuações grandes em seus extremos quando comparados com a região central. Isso ocorre porque os eventos relacionados aos extremos são pouco frequentes e, portanto, muito sujeitos a flutuações. Flutuações essas que teremos várias oportunidades de nos depararmos no decorrer do nosso trabalho.

Agora não apresentaremos exemplos sobre a distribuição de Laplace, pois haverá muitos nos capítulos subseqüentes.

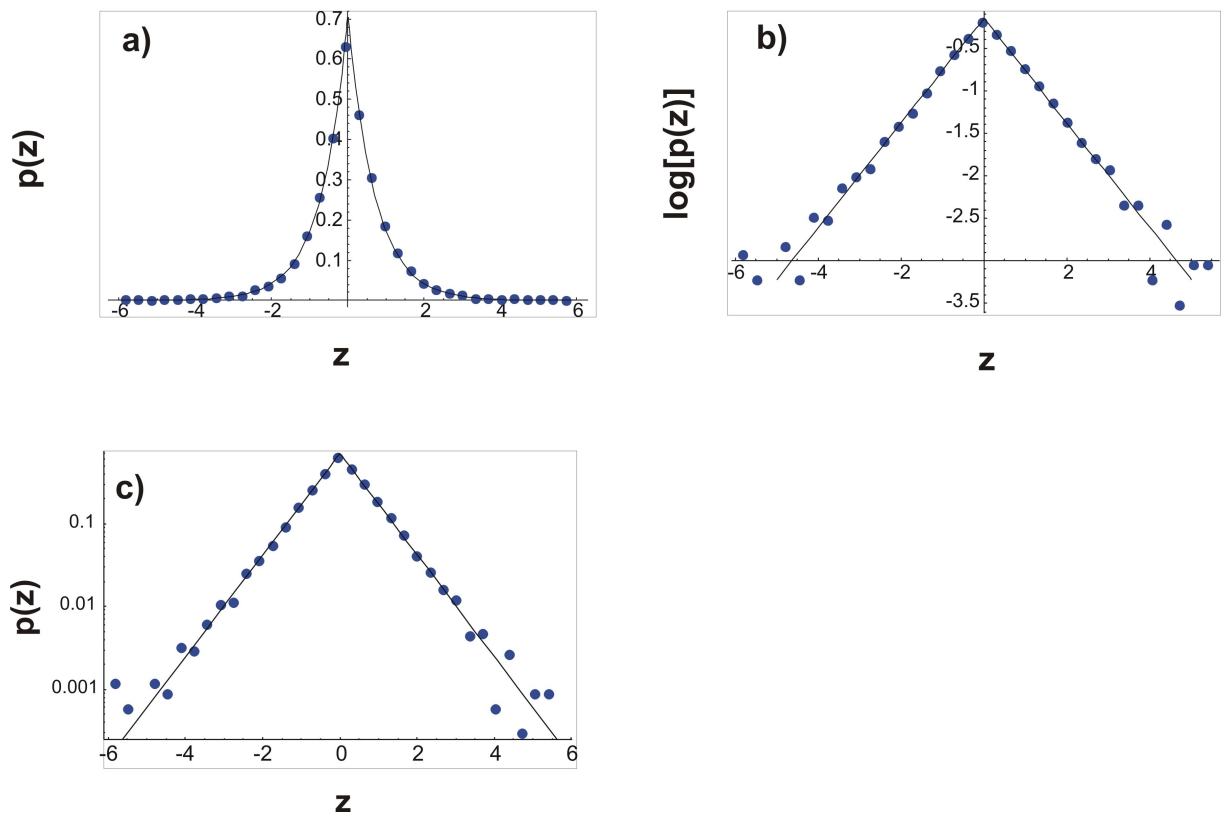


Figura 1.22: Histograma para dados aleatórios simulados seguindo uma distribuição de Laplace com desvio padrão unitário e média nula. Em a), exibimos o gráfico normal, em b), o gráfico do logaritmo da distribuição e, em c), o mono-log. A linha contínua representa a distribuição e os pontos o histograma.

Capítulo 2

Séries Temporais

Neste capítulo, apresentamos conceitos fundamentais para o desenvolvimento de nossos estudos, indo desde o conceito de séries temporais até taxas de crescimentos. Também será exposto o *Modelo de Gibrat* e um outro modelo estocástico proposto na referência [13], que consistem em modelos que estão relacionados aos tipos de análises que serão empregados em nosso trabalho.

2.1 Introdução às Séries Temporais

Este trabalho baseia-se fundamentalmente em analisar séries temporais, que por sua vez, trazem importantes informações sobre a dinâmica do sistema a que elas se referem. Uma série desse tipo é constituída por uma seqüência dos valores de uma determinada grandeza ordenados temporalmente, por exemplo, em valores diários, mensais, bimestrais, anuais, etc.

As séries temporais podem ocorrer de diferentes formas. Por um lado, podemos conhecer informações a respeito de poucas partes do sistema para muitos intervalos de tempo e assim temos poucas séries, mas cada uma com muitos termos. Em particular, poderíamos ter apenas uma série. Em um outro extremo, teríamos muitas séries pequenas, cada uma correspondendo a uma parte do sistema. De forma mais geral, poderíamos representar as séries temporais da seguinte forma:

$$\begin{array}{cccc} z_1(1) & z_1(2) & \dots & z_1(N) \\ z_2(1) & z_2(2) & \dots & z_2(N) \\ & & \vdots & \\ z_n(1) & z_n(2) & \dots & z_n(N) . \end{array} \quad (2.1)$$

Note que as horizontais são as séries temporais, próprias de cada uma das n partes do sistema. Porém, as verticais também podem ser consideradas séries (não temporais), pois contém diversas partes do sistema em um dado momento. Uma ilustração de (2.1) seriam as linhas horizontais corresponderem aos PIB's de n países em N anos, uma linha vertical qualquer é a série (não temporal) dos PIB's dos n países em um dado ano.

Dada uma série temporal, podemos construir outras séries derivadas, por exemplo, relacionadas à flutuação dos valores da série original. Uma possibilidade seria considerar as variações entre sucessivos valores de uma série temporal ou, de uma maneira mais geral,

$$Q_s(t) = z(t + s) - z(t) , \quad (2.2)$$

em que $z(t + s)$ e $z(t)$ são valores de z nos instantes $t + s$ e t , respectivamente.

Tal definição é encontrada na análise de vários sistemas, por exemplo, na caminhada humana [27]. Não se espera que o tamanho dos sucessivos passos venham a aumentar ou diminuir sistematicamente, entretanto o comprimento de cada passo não é, em geral, igual ao anterior e nem ao posterior. Outros exemplos são estudos de índices econômicos [15, 16], temperatura ambiente [64] e batimentos cardíacos [26].

Certos sistemas apresentam regularidades em seu crescimento, como a do tipo exponencial. Em tal tipo, para intervalos de tempo iguais e consecutivos, há o aumento por um mesmo fator multiplicativo. Entretanto, em alguns sistemas o crescimento é similar ao exponencial, porém, o fator multiplicativo geralmente muda a cada instante considerado, como é o caso da poupança mês a mês. Nesses sistemas, encontramos um outro tipo de flutuações, e que será de particular interesse em nosso estudo.

Visando entender alguns aspectos desse crescimento, de aparente aleatoriedade, começamos colocando em evidência o comportamento puramente exponencial. Seja $z(1)$ o valor inicial da variável z , que cresce exponencialmente, e \tilde{R} o fator de crescimento (ou decrescimento), para um intervalo de tempo considerado. Assim, a evolução de z poderá ser prevista, e no n -ésimo intervalo considerado será dado por $z(n) = \tilde{R}^{n-1} z(1)$, pois $z(2) = \tilde{R} z(1)$, $z(3) = \tilde{R}^2 z(1)$ etc. Por outro lado, como foi dito anteriormente, flutuações da dinâmica exponencial podem ocorrer. Para ilustrar tal fato, consideramos o caso da poupança em que a taxa de crescimento depende do mês considerado. Chamando de $\tilde{R}_1(t)$ a taxa de crescimento no t -ésimo mês, então o valor de z após os n meses será dado por

$$\begin{aligned} z(2) &= \tilde{R}_1(1) z(1) \\ z(3) &= \tilde{R}_1(2) z(2) = \tilde{R}_1(2) \tilde{R}_1(1) z(1) \\ &\vdots \\ z(n) &= \tilde{R}_1(n-1) z(n-1) \\ &= \tilde{R}_1(n-1) \dots \tilde{R}_1(2) \tilde{R}_1(1) z(1), \end{aligned} \quad (2.3)$$

sendo $z(1)$ o capital inicialmente investido. Por sua vez, se tivermos os valores de z nos diversos meses, podemos conhecer a taxa de crescimento em um determinado mês por meio da relação

$$\tilde{R}_1(t) = \frac{z(t+1)}{z(t)} . \quad (2.4)$$

De uma maneira geral, em sistemas nos quais encontramos crescimento (decréscimo), é comum considerar flutuações, para termos não consecutivos, do tipo

$$R_s(t) = \ln \tilde{R}_s(t) = \ln \frac{z(t+s)}{z(t)} . \quad (2.5)$$

Objetivando simplificar a notação ao longo do nosso texto, passaremos a empregar R para representar a taxa de crescimento logarítmica, R_1 ,

$$R(i) = \ln \frac{z(i+1)}{z(i)} . \quad (2.6)$$

Utilizaremos também o logaritmo na base 10. Assim, a equação (2.4) passará a ser escrita como

$$R(t) = \log \frac{z(t+1)}{z(t)}. \quad (2.7)$$

Gostaríamos de ressaltar que a definição 2.6 tem sido utilizada, por exemplo, no estudo da dinâmica de populações de pássaros [24], tamanho de firmas [17, 18, 19], produto interno bruto dos países [22], pesquisas em universidades [23] e produção científica [9].

2.2 Análise das Flutuações

Em investigações que levam em consideração as taxas de crescimento, definida via equação (2.7), é comum encontrarmos distribuições de probabilidades exponenciais

$$p(R) = \frac{1}{\sqrt{2}\sigma} \exp\left(-\frac{\sqrt{2}|R - \mu|}{\sigma}\right), \quad (2.8)$$

em que $\mu = \bar{R}$ é o valor médio de R e σ o desvio padrão das flutuações. Sua representação num gráfico mono-log se assemelha a uma tenda, por isso o nome *tent shape* (forma de tenda), veja a figura 2.1. Esse tipo de distribuição, assim como uma lei de potência, decai bem mais lentamente que a distribuição gaussiana, sugerindo a presença de flutuações maiores que as esperadas para o caso de uma distribuição gaussiana.

Uma importante propriedade nesse tipo de estudo, comumente observada para vários sistemas, é o fato de que o desvio padrão apresentado nas distribuições das flutuações possuem dependência do tipo lei de potência, com o “tamanho” S do subsistema, isto é, $\sigma(S) \propto S^{-\beta}$, situação que exemplificamos no próximo capítulo.

Isso é compatível com o fato de que os histogramas tenham a mesma forma, *tent-shape*, porém com larguras diferentes. Quanto menores forem os PIB's, por exemplo, mais largo será o gráfico em 2.1a. Como poderíamos esperar, uma grande firma, em geral, não cresce ou diminui muito em um dado ano. É possível, no entanto, normalizar tais distribuições, fazendo com que os dados colapsem em uma única curva. Em [21], a normalização foi feita de tal forma que, $p(r) \equiv \sigma p(R)$ e $r \equiv (R - \mu)/\sigma$, assim, os dados relativos ao tamanho das firmas colapsam em uma única curva, $p(r) \propto \exp(-\sqrt{2}|r|)$, veja a figura 2.1b.

O que é notável sobre a equação (2.8) é que ela, com boa aproximação, governa as taxas de crescimento de um conjunto diverso de firmas. Sua abrangência não se limita aos tamanhos, mas também ao que produzem e a quem produzem. Não há, nas teorias convencionais de economia, sugestão de que diferentes setores, como indústria de automóveis, de produtos farmacêuticos, ou ainda de papéis, possam ser governados pelas mesmas leis de crescimento [17]. Por sua vez, os resultados encontrados apontam para um tipo de *universalidade*, em que diferentes sistemas são caracterizados pelas mesmas leis, independentes de detalhes “microscópicos” [17, 23]. Na figura 2.2, vemos a concordância entre os dados relativos ao tamanho das vendas e ao número de empregados, embora também tenha sido igualmente consistente o preço dos produtos vendidos, o patrimônio das firmas e a estrutura (construções e equipamentos).

Com a idéia de buscar um comportamento universal, comparou-se o crescimento de diferentes sistemas verificando um ajuste bastante satisfatório. Em [17], apresentaram a mesma dinâmica de crescimento as vendas, o número de empregados, o preço dos bens vendidos, o patrimônio das firmas e as respectivas estruturas. Já em [23], a mesma forma

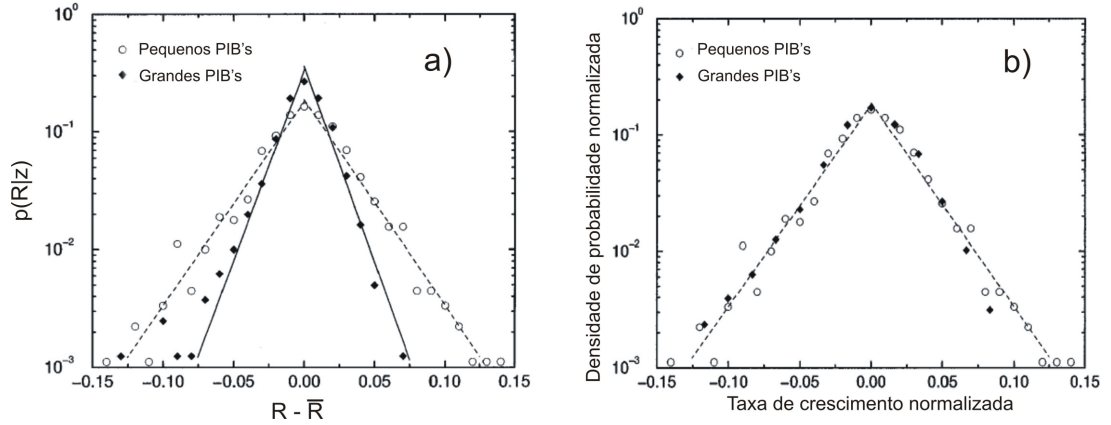


Figura 2.1: Figura retirada e adaptada da referência [21], apresenta, em a), a distribuição de probabilidades condicional para pequenos e grandes PIB's. Nota-se aqui que embora apresentem a mesma forma funcional, as larguras das distribuições são diferentes. As linhas tracejada e contínua correspondem a equação (2.8). Em b), empregada a taxa de crescimento normalizada $r = (R - \mu)/\sigma$, vê-se as distribuições de probabilidades colapsando em uma única curva, ainda com a mesma forma. A linha tracejada corresponde a uma função a $\exp(-\sqrt{2}|r|)/\sqrt{2}$.

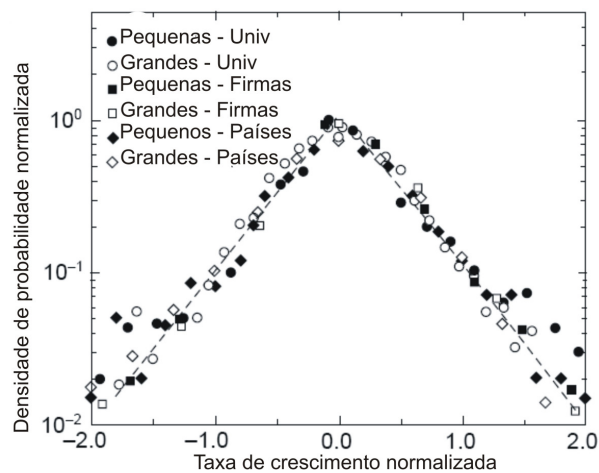


Figura 2.2: Figura retirada da referência [23]. Nota-se que, após a normalização, dados relativos a diferentes campos colapsam em uma única curva.

foi encontrada para o crescimento de universidades, de firmas e até mesmo do PIB dos países.

É interessante ressaltar que o conceito de universalidade teve origem quando se percebeu que diversas grandezas termodinâmicas, tais como calor específico, compressibilidade e susceptibilidade magnética, apresentavam um comportamento universal próximos a um ponto crítico de uma transição de fase de segunda ordem, caracterizados pelos mesmos valores de *expoentes críticos* [62], quando várias classes de sistemas foram investigados.

2.3 Modelo de Gibrat e generalização

Em se tratando de séries e, em particular, as investigadas neste trabalho, um outro ponto importante é o modelo de Gibrat. Proposto na primeira metade do séc. XX [63], objetivou explicar a dinâmica de crescimento de firmas e baseou-se nas seguintes hipóteses: *i*) a taxa de crescimento de uma firma é independente de seu tamanho; *ii*) as taxas de crescimento sucessivas de uma firma não apresentam correlações temporais; e *iii*) as firmas não interagem. Na forma matemática, este modelo pode ser expresso pelo processo estocástico multiplicativo

$$S(t+1) - S(t) = \epsilon(t)S(t) , \quad (2.9)$$

sendo $S(t+1)$ e $S(t)$ os tamanhos de uma firma nos tempos $t+1$ e t , e $\epsilon(t)$ é um número aleatório não correlacionado seguindo uma distribuição de probabilidades Gaussiana com variância muito menor que a unidade.

Em diversas investigações, é de interesse a análise da distribuição de tamanhos para o sistema estudado. Esse “tamanho” pode ser, entre outras coisas, o faturamento de uma firma [17, 18, 19], o valor do PIB [21, 22] de um país, o gasto em pesquisa e desenvolvimento por uma universidade [23] e o número de artigos publicados [9]. Nesse contexto, é comum a referência ao modelo de Gibrat [63], em que a distribuição dos tamanhos se aproximam de uma lognormal. Entretanto, nestes sistemas em que a distribuição dos tamanhos segue um padrão lognormal, outras propriedades não são caracterizadas pelo modelo de Gibrat.

O modelo proposto por Gibrat tem sido utilizado numa análise inicial de estudos de dinâmicas de crescimentos. É comum encontrar trabalhos fazendo referência ao modelo, mesmo que o intuito seja afirmar que os resultados não são concordantes.

Visando ilustrar o modelo de Gibrat, fizemos uma simulação computacional, onde consideramos 2×10^4 equações do tipo (2.9) partindo, todas elas, de $S(1) = 1$ e com $\epsilon(t)$ seguindo uma distribuição normal, com $\sigma = 0,001$ e $\mu = 0$. Após 10^5 iterações dessas equações independentes, fizemos um histograma com os $S(100000)$'s. Neste trabalho, a forma mais comum de representar graficamente a distribuição dos tamanhos será via um histograma do logaritmo da distribuição pelo logaritmo de S . Encontramos o resultado desta simulação na figura 2.3.

Deve ser notado que a figura 2.3 mostra que os resultados são aproximados por uma lognormal. Essa aproximação será tanto melhor quanto mais iterações houver, pois somente no limite de infinitas iterações ocorrerá exatamente uma distribuição lognormal, conforme discutimos no apêndice B.

Um outro modelo de crescimento foi edificado recentemente [13] e tem o mérito de reproduzir algumas características do sistema que estudamos, pontos em que Gibrat falha.

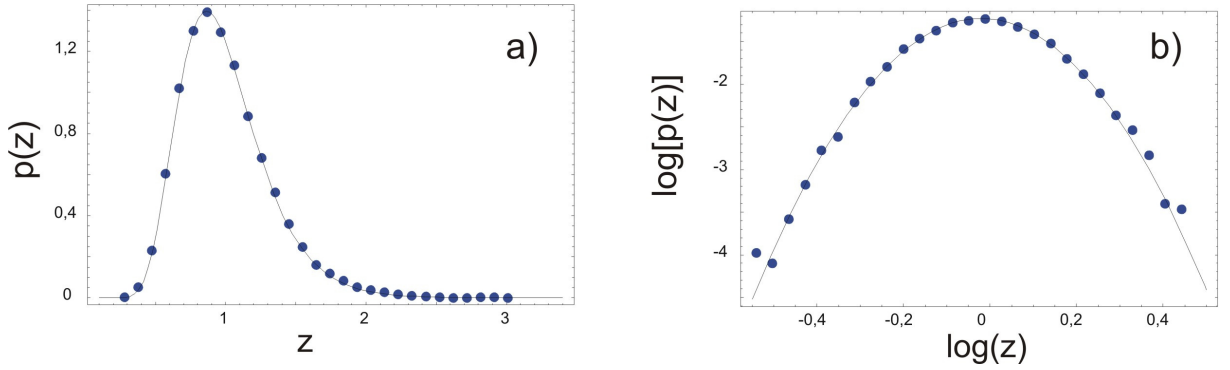


Figura 2.3: Histograma obtido a partir da simulação do modelo de Gibrat com 2×10^4 partes e 10^5 iteradas. Em a), é exibido o gráfico usual da lognormal (linha contínua) e os pontos compõem o histograma dos dados simulados. Em b), o logaritmo da lognormal pelo logaritmo de z é representada por uma parábola (linha contínua) e os pontos dizem respeito à simulação.

Seguindo a generalização do modelo de Gibrat proposta na referência [13], consideremos que a grandeza $F(t)$ é dada por um processo multiplicativo não linear que evolui no tempo segundo a equação

$$F(t+1) - F(t) = \lambda(t)[F(t)]^k, \quad (2.10)$$

com $k > 0$ e $F \geq F_m$, sendo F_m uma constante positiva. Suponha também que $\lambda(t)$ seja dado por um processo multiplicativo do tipo

$$\lambda(t) = [b_0 + b_1 \lambda(t-1) \lambda(t)], \quad (2.11)$$

com $b_0 \ll 1$ e b_1 sendo constante positivas. O termo $\lambda(t)$ refere-se a um número aleatório cuja distribuição é gaussiana, com média zero e variância unitária. O processo estocástico definido pelas equações 2.10 e 2.11 recupera o modelo de Gibrat (na ausência do corte F_m) quando $k = 1$ e $b_1 = 0$.

Verificamos que esta generalização do modelo de Gibrat é consistente com a equação $\sigma(F) \sim F^{-L}$, com $L > 0$, indicando que o desvio padrão das taxas de crescimento são dependentes do tamanho. Verificamos também que o valor médio das taxas de crescimento exibe um comportamento geral independente do tamanho, flutuando em torno de zero, principalmente para valores de F distantes da barreira refletora definida por F_m .

A dinâmica do modelo é controlada basicamente por dois parâmetros: k e b_1 . O efeito destes parâmetros nas previsões do modelo está ilustrado na figura (2.4), onde vemos resultados de simulações para diferentes valores de k e b_1 .

O parâmetro b_1 proporciona um efeito de memória. Este parâmetro é o principal responsável pela forma da distribuição das taxas de crescimento, gerando distribuições Gaussianas se $b_1 = 0$ e não-gaussianas se $b_1 \neq 0$. Quanto maior o valor de $0 < b_1 \leq 1$, mais a distribuição se desvia da Gaussiana figura (2.4a). Por outro lado, a forma da distribuição das taxas de crescimento é praticamente independente de k (ver figura 2.4b).

O parâmetro k reflete o fato de que as taxas de crescimento dependem do tamanho F . Se $k < 1$ as flutuações nas taxas de crescimento são amplificadas para F pequeno e amortecidas para F grande. Por outro lado, se $k > 1$ as flutuações são amortecidas para F pequeno e amplificadas para F grande. Deste modo, o parâmetro k governa a

dependência do desvio padrão σ com F em $\sigma(F) \simeq F^{-L}$ (figura 2.4c). As simulações realizadas em [13] indicam que este parâmetro pode ser relacionado com o expoente L pela relação $L \simeq 1 - k$. Por sua vez, o expoente L não apresenta dependência significativa com b_1 (figura 2.4d).

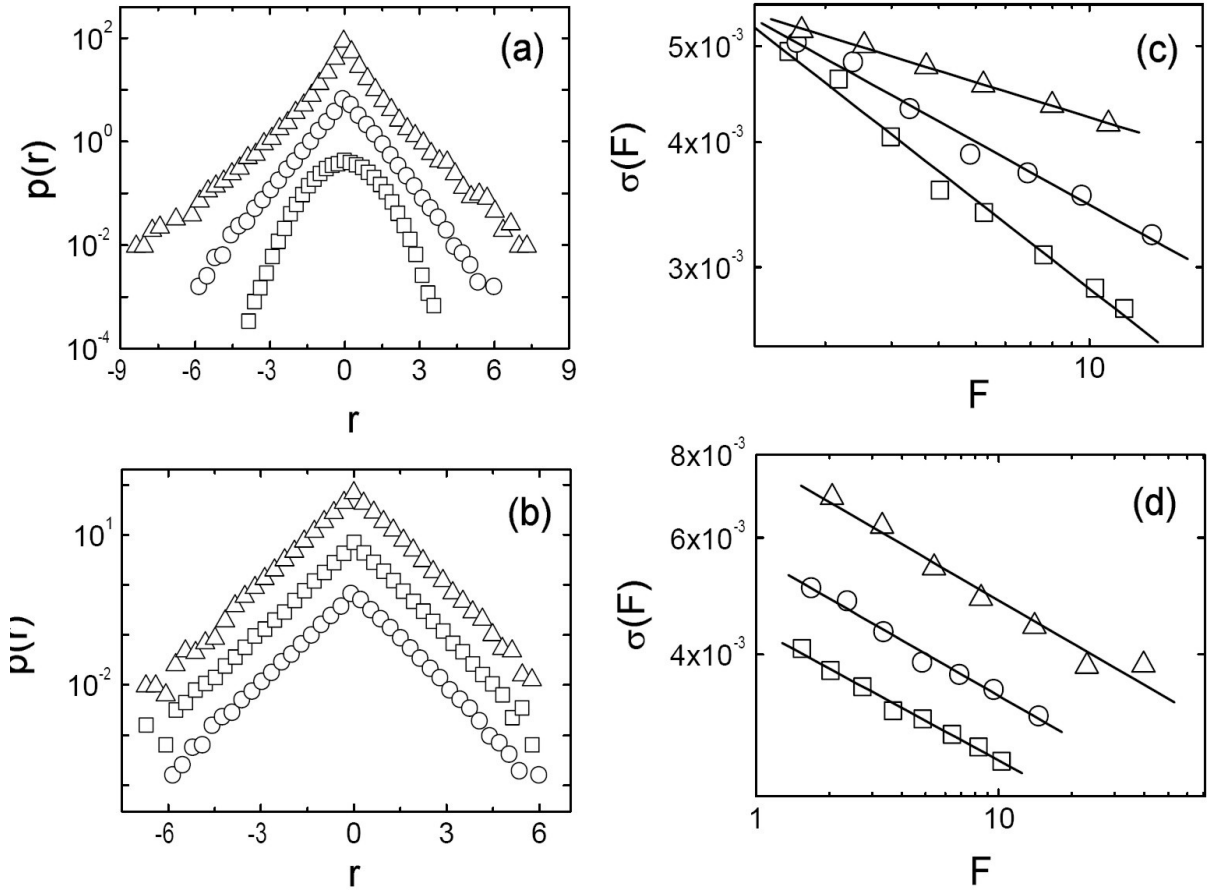


Figura 2.4: Figura retirada da referência [13]. Realização típica do modelo definido pelas equações 2.10 e 2.11 com 1000 elementos de um sistema, todos com valor inicial de $F_0 = 1$, com $F \geq F_m = 1$. Fixou-se $b_0 = 0,01$ e variou-se os valores de b_1 e k . Em cada caso foi considerado 100 passos depois de um regime inicial. (a) Densidade de probabilidade $p(r)$, para $k = 1$ e $b_1 = 0$ (quadrados), $b_1 = 0,6$ (círculos) e $b_1 = 0,9$ (triângulos). (b) Densidade de probabilidade das taxas de crescimento normalizadas, $p(r)$, para $b_1 = 0,6$ e $k = 1$ (círculos), $k = 0,8$ (quadrados) e $k = 0,5$ (triângulos). Em (a) e (b) as curvas foram deslocadas na vertical para melhor visualização. (c) Desvio padrão das taxas de crescimento, $\sigma(F)$, para $b_1 = 0,6$ e $k = 0,7$ (quadrados), $k = 0,8$ (círculos) e $k = 0,9$ (triângulos). As linhas retas são ajustes lineares representando leis de potência com expoente $\beta \simeq 0,3$, $\beta \simeq 0,2$ e $\beta \simeq 0,1$, respectivamente. (d) Desvio padrão das taxas de crescimento, $\sigma(F)$, para $k = 0,8$ e $b_1 = 0$ (quadrados), $b_1 = 0,6$ (círculos) e $b_1 = 0,8$ (triângulos). As linhas retas representam leis de potência com expoente $\beta = 0,2$.

Capítulo 3

Aplicações Recentes

Neste capítulo, elencamos algumas recentes investigações baseadas em séries temporais e mais voltadas ao ponto central de nosso estudo. As figuras utilizadas neste capítulo foram retiradas de alguns artigos a que nos referimos, e por isso, a notação poderá não coincidir com a que vem sendo utilizada e, quando for o caso, explicitaremos a relação entre as grandezas deste trabalho e as presentes nos artigos.

3.1 Citações de Periódicos Científicos

Como ilustração de uma recente aplicação dos métodos expostos no capítulo anterior, citamos a dinâmica de crescimento dos fatores de impacto de periódicos científicos [12]. Foi utilizado um conjunto de dados que inclui o fator de impacto de mais de 5000 periódicos em vários anos. Inclui também o fator de impacto de 2910 periódicos no período de 13 anos: de 1992 a 2004.

O fator de impacto de periódicos científicos foi proposto na década de 1950 [73, 74]. É definido como a razão entre o número de citações que um periódico recebe no ano corrente (devido a artigos publicados nos dois anos anteriores) pelo número de artigos publicados nos dois anos anteriores. Assim, para calcular o fator de impacto de um periódico relativo a 2006, dividimos o total de citações que ele recebeu neste ano referente aos artigos publicados em 2004 e 2005, pelo número total de artigos publicados em 2004 e 2005. O fator de impacto é calculado anualmente pelo *Institute for Scientific Information* (ISI) e publicado no *Journal Citation Report* (JCR) [75].

Inicialmente foi calculada a distribuição do logaritmo do fator de impacto para 5912 periódicos no ano de 2004, ilustrado na figura 3.1a. Vemos a distribuição de probabilidades $p(\log F)$ em comparação com a gaussiana. O ajuste se dá de forma satisfatória apenas para valores intermediários de F , indicando que $p(F)$ somente poderá ser ajustado por uma distribuição lognormal para valores intermediários de F . Além disso, a figura 3.1b mostra que a distribuição $p(F)$ apresenta um comportamento do tipo lei de potência para pequenos e grandes valores de F . Quantitativamente,

$$p(F) \sim F^{-\alpha} , \quad (3.1)$$

com $\alpha \simeq 2,7$ para F grande e $\alpha \simeq -0,4$ para F pequeno.

Visto que $p(F)$ apresenta diferentes comportamentos para pequenos, intermediários e grandes valores de F , uma distribuição mais geral foi utilizada para descrever tais

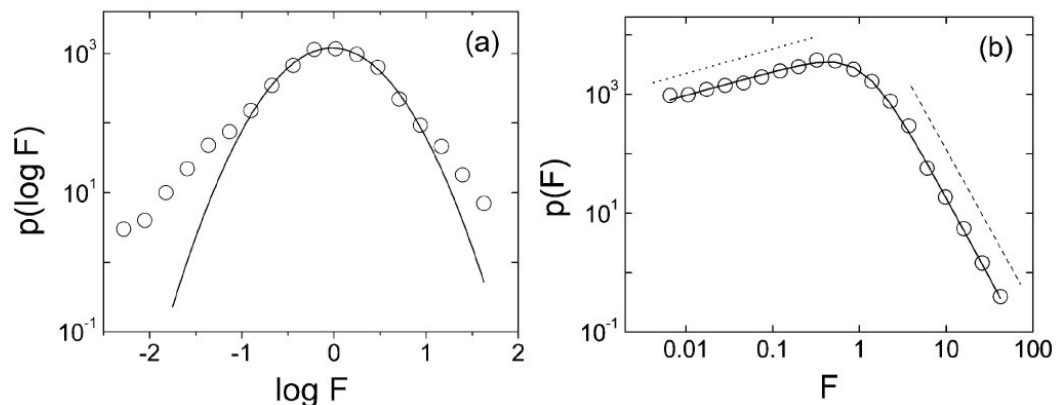


Figura 3.1: Figura retirada da referência [12]. Em a), vê-se a densidade de probabilidade $p(\log F)$ para os referidos periódicos (círculos), em que F é o fator de impacto. A linha sólida representa uma distribuição gaussiana na variável $\log F$ (portanto, uma lognormal em F). Em b), evidencia-se que a distribuição $p(F)$ tem um comportamento do tipo lei de potência para pequenos e grandes valores de F . A linha pontilhada a esquerda representa uma função $p(F) \propto F^{0,4}$, a pontilhada a direita uma função $p(F) \propto F^{-2,7}$ e a linha sólida $p(F) \propto F^{0,4}(1,3 + F^2)^{-1,55}$.

comportamentos. Na figura 3.1b, a linha sólida representa a função $p(F) \propto F^\gamma(a + F^2)^{-\alpha}$, com $\gamma = 0,4$, $a = 1,3$ e $\alpha = 1,55$.

O próximo passo foi investigar a dinâmica das citações dos periódicos, estudando as flutuações anuais dos fatores de impacto. Foram selecionados, a partir dos dados, periódicos que apresentavam fatores de impacto de todo o período considerado (1992-2004), um total de 2910 periódicos. Para cada periódico selecionado, foi calculada a taxa de crescimento, como definido em (2.5),

$$R(t) = \log \left[\frac{F(t+1)}{F(t)} \right], \quad (3.2)$$

com $F(t)$ e $F(t+1)$ sendo os fatores de impacto nos anos t e $t+1$, respectivamente. Tem-se, então, um conjunto de 2910 séries, com 12 termos cada uma. Para determinar a distribuição das taxas de crescimento, agrupam-se tais séries, pois permitem construir um histograma e assim determinar a forma da distribuição.

Os periódicos foram divididos em grupos de acordo com os respectivos valores iniciais de F . Para cada grupo, calculou-se a densidade de probabilidade condicional, $p(R|F)$, das taxas de crescimento do fator de impacto. A figura 3.2a mostra a densidade de probabilidade condicional $p(R|F)$ calculada para três grupos de periódicos. Observe que a distribuição exibe um comportamento que pode ser bem aproximado por uma exponencial, dada por

$$p(R|F) = \frac{1}{\sqrt{2}\sigma} \exp \left(-\frac{\sqrt{2}|R - \mu|}{\sigma} \right), \quad (3.3)$$

em que μ e σ são a média e o desvio padrão de R , respectivamente, calculados dentro de cada grupo. O resultado apresenta a mesma forma funcional, porém, quanto maior o valor inicial do grupo mais estreita é a largura da curva (desvio padrão). Utilizando os mesmos

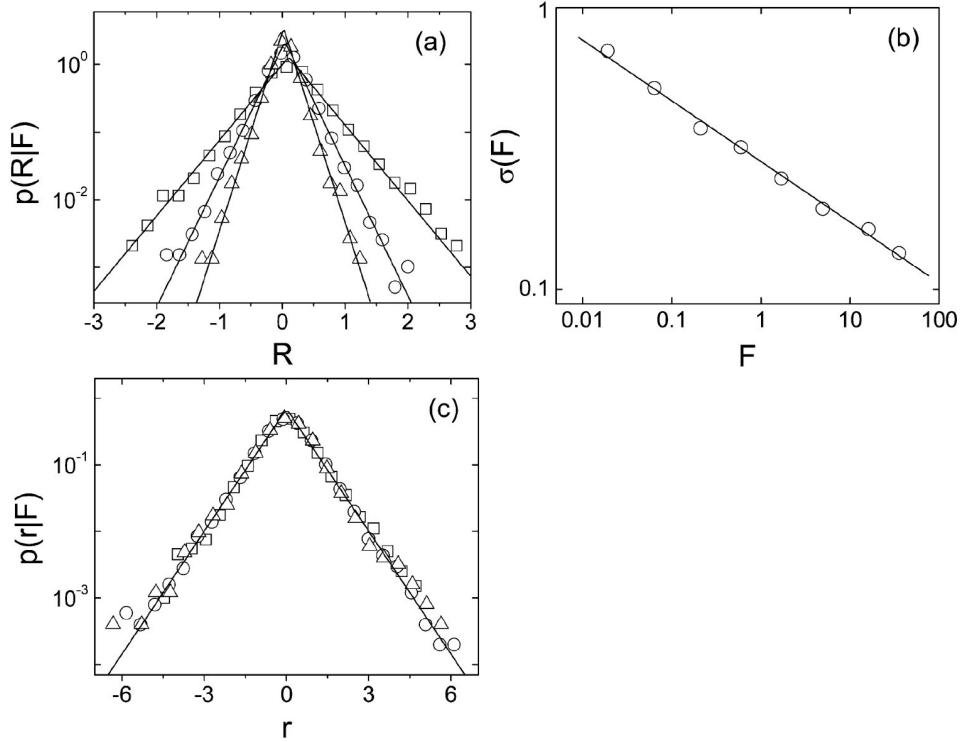


Figura 3.2: Figura retirada da referênciã [12]. Em a), tem-se as distribuições de probabilidades condicionais $p(R|F)$, para três grupos distintos, usando como critério o tamanho do fator de impacto, $0, 1 < F < 0, 2$ (quadrados), $0, 3 < F < 0, 7$ (círculos) e $1, 3 < F < 2$ (triângulos). A mesma forma funcional é verificada, porém com diferentes larguras. Em b), o desvio padrão decresce com o valor do fator de impacto F , justificando as diferenças de largura em a). Em c), a normalização faz com que haja o colapso das distribuições de a) em uma única curva.

grupos de periódicos, foi calculada a distribuição das taxas de crescimento normalizadas

$$r(t) = \frac{R(t) - \mu}{\sigma}, \quad (3.4)$$

assim, as três curvas colapsam em uma única, como mostrado na figura 3.2c, é descrita por

$$p(r|F) = \frac{1}{\sqrt{2}} \exp(-\sqrt{2}|r|). \quad (3.5)$$

Para investigar em detalhes a dependência do desvio padrão com o valor inicial do fator de impacto, o grupo de periódicos foi dividido em oito e, então, calculou-se o desvio padrão em função do valor médio do fator de impacto inicial. O resultado é mostrado na figura 3.2b, que sugere que o desvio padrão segue uma lei de potência:

$$\sigma(F) \propto F^{-\beta}, \quad (3.6)$$

com $\beta \simeq 0, 22$.

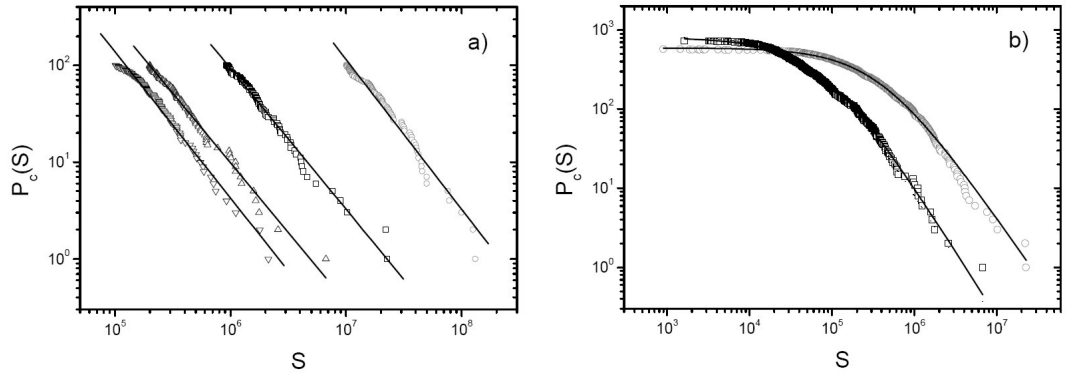


Figura 3.3: Figura retirada da referência [25]. Em a), ilustra-se a distribuição dos 100 maiores jornais dos EUA em 2003 (triângulos para baixo), 100 maiores revistas do RU em 2004 (triângulos para cima), 100 maiores revistas dos EUA (quadrados) e 100 maiores jornais do RU (círculos), com o expoente $\alpha \simeq 1,5$. Em b), em ambos os conjuntos de dados, $P_C(S)$ apresenta um comportamento assintótico do tipo lei de potência.

3.2 Circulação de Revistas e Jornais

Continuando com os exemplos das aplicações dos métodos estatísticos que utilizamos neste trabalho, consideramos agora as propriedades estatísticas da circulação de revistas e jornais [25]. O dados analisados foram retirados do *Audit Bureau of Circulations* (ABC).

Circulação, número de cópias produzidas, distribuídas e vendidas de um volume específico, pode ser vista como uma medida do número de leitores de uma dada revista ou jornal. Além do mais, a circulação é um importante parâmetro para medir a receita destas empresas, e de certa forma, está relacionada com o tamanho das mesmas.

A primeira etapa foi analisar a cauda da distribuição da circulação anual das revistas e jornais. Aqui a cauda refere-se às 100 maiores revistas e jornais de diferentes conjuntos de dados. Para tal, foi utilizada a distribuição acumulada, figura 3.3a, que apresenta a forma de uma lei de potência, para os diferentes dados.

A função correspondente à distribuição acumulada é definida por

$$P_C(S) = \int_S^{\infty} P(s) ds . \quad (3.7)$$

Assim, quando $P(S) \propto S^{-1-\alpha}$, obtem-se $P_C(S) \propto S^{-\alpha}$, e α pode ser estimado pela inclinação da reta no gráfico log-log.

Analisou-se também um conjunto de dados mais amplo, com 570 revistas dos Estados Unidos da América (EUA) e 727 do Reino Unido (RU), em que pode-se notar que o comportamento assintótico é também do tipo lei de potência. Novamente, uma distribuição mais geral foi empregada. A lei de Zipt-Mandelbrot [76] apresenta um comportamento assintótico do tipo lei de potência e foi utilizada para ajuste dos dados. Pode-se representá-la por

$$P(S) = \frac{d}{(c + S)^\alpha} , \quad (3.8)$$

com d , c e α sendo constantes. Veja a figura 3.3b.

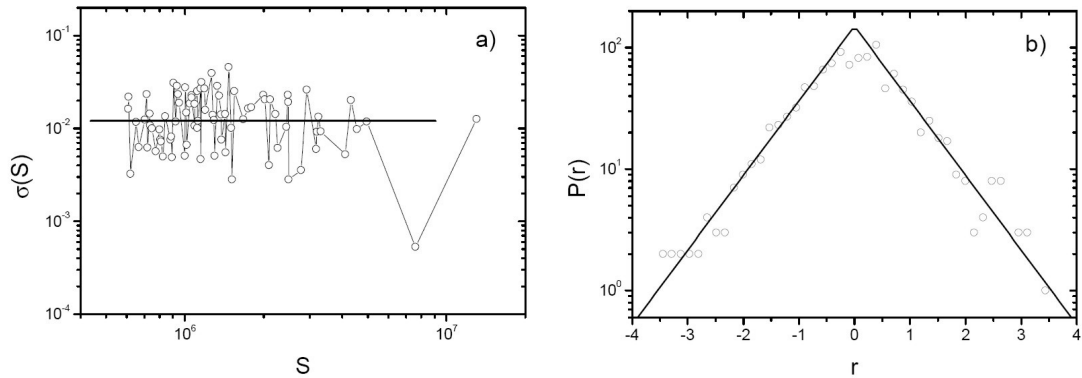


Figura 3.4: Figura retirada da referência [25]. Tem-se o desvio padrão das taxas de crescimento em função da circulação, em a), e a distribuição das taxas de crescimento normalizada para revistas com grande circulação, em b).

Também investigou-se a dinâmica de crescimento da circulação das revistas. Para tal, usou-se o logaritmo da taxa de crescimento (veja a equação 2.6):

$$R(t) = \ln \frac{S(t)}{S(t-1)}, \quad (3.9)$$

em que $S(t)$ e $S(t-1)$ correspondem à circulação nos instantes t e $t-1$, respectivamente. Utilizaram-se 82 revistas norte-americanas, que apresentavam dados semestrais completos no período (1996-2003) e calcularam-se as taxas de crescimento R .

Posteriormente, foi calculado o desvio padrão, σ , das taxas de crescimento para cada uma das revistas consideradas. A figura 3.4a sugere que σ seja independente de S . O resultado parece ir em contradição ao esperado, visto que o desvio padrão dos fatores de impacto decai com o tamanho das revistas segundo uma lei de potência com expoente $\beta \simeq 0,22$ (seção 3.1), o desvio padrão dos gastos também decai com o tamanho das universidades como uma lei de potência de expoente $\beta = 0,25$ (seção 3.3). Deve-se enfatizar que essa aparente contradição ocorre em virtude de ter sido considerado apenas um conjunto pequeno, as revistas com maior circulação.

Utilizando a taxa de crescimento normalizada, as distribuições das taxas de crescimento são colapsadas em uma única curva, na forma de “tenda”, como mostrado na figura 3.4b. A taxa de crescimento normalizada r , como no exemplo da seção anterior, é expressa por

$$r(t) = \frac{R(t) - \mu}{\sigma}, \quad (3.10)$$

em que μ e σ são, respectivamente, a média e o desvio padrão das taxas de crescimento para cada revista considerada. A distribuição normalizada é consistente com a distribuição exponencial $p(r) \propto \exp(-\sqrt{2}|r|)$.

3.3 Pesquisas em Universidades

Na referência [23], encontra-se novamente o estudo de propriedades estatísticas, agora em relação às atividades de pesquisa em universidades norte-americanas. Foi analisado um conjunto de cinco diferentes bancos de dados, dos quais o maior refere-se ao *National*

Science Foundation com informações a respeito dos gastos com pesquisa e desenvolvimento (P&D) pelos centros de ciência e engenharia. Ao todo, foram investigados os gastos de 719 universidades em um período de 17 anos (1979-1995).

Os resultados ainda foram comparados com estudos anteriores sobre a dinâmica de crescimento de firmas [17] e do PIB dos países [21]. As distribuições das taxas de crescimento apresentam uma forma universal e não dependem do tamanho da universidade e foi encontrado também que as larguras das distribuições decaem com o tamanho das universidades como uma lei de potência. A comparação mostrou que os resultados são similares aos resultados do crescimento de firmas e, assim, são consistentes com a hipótese de que a dinâmica de crescimento de organizações complexas são governadas por mecanismos universais [21].

A taxa de crescimento anual dos gastos com P&D, $R(t)$, é definida como em (2.7),

$$R(t) = \log \left[\frac{S(t+1)}{S(t)} \right], \quad (3.11)$$

sendo $S(t+1)$ e $S(t)$ os gastos nos anos $t+1$ e t , respectivamente. Espera-se que as propriedades estatísticas da taxa de crescimento dependam de S , assim como é natural que as flutuações diminuam conforme aumentam-se os gastos.

As universidades foram divididas em grupos, de acordo com o valor inicial de S . A figura 3.5a representa a distribuição do logaritmo dos gastos e a figura 3.5b a probabilidade condicional $p(R|S)$ da taxa de crescimento anual. Note que, quanto maiores os gastos, mais estreita é a curva. Tais curvas são bem descritas por distribuições exponenciais, como nos exemplos anteriores,

$$p(R|S) = \frac{1}{\sqrt{2}\sigma} \exp \left(\frac{-\sqrt{2}|R - \mu|}{\sigma} \right), \quad (3.12)$$

em que μ e σ são, respectivamente, a média e o desvio padrão de R para cada grupo considerado.

É evidenciado, por meio da figura 3.5c, a dependência do desvio padrão com o valor inicial de S . Mais precisamente, nota-se que o desvio padrão decai segundo uma lei de potência,

$$\sigma(S) \propto S^{-\beta}, \quad (3.13)$$

com expoente $\beta = 0,25$.

As distribuições de probabilidades condicionais normalizadas apresentam mesma forma. A figura 3.5d mostra que os três grupos são agregados em uma única curva,

$$p(r|S) = \frac{1}{\sqrt{2}} \exp(-\sqrt{2} r), \quad (3.14)$$

com $r = |R - \mu|/\sigma$, e novamente, μ é a média e σ o desvio padrão nos grupos considerados.

Estes resultados ainda foram comparados com outros bancos de dados e o mesmo comportamento foi encontrado. Tais bancos de dados referem-se ao número de artigos publicados por 112 universidades norte-americanas, ao número de patentes destas universidades e ainda a fundos de pesquisa de universidades canadenses e inglesas.

As comparações mostram que o comportamento do desvio padrão é idêntico, decai com o tamanho da grandeza considerada, segundo uma lei de potência com expoente

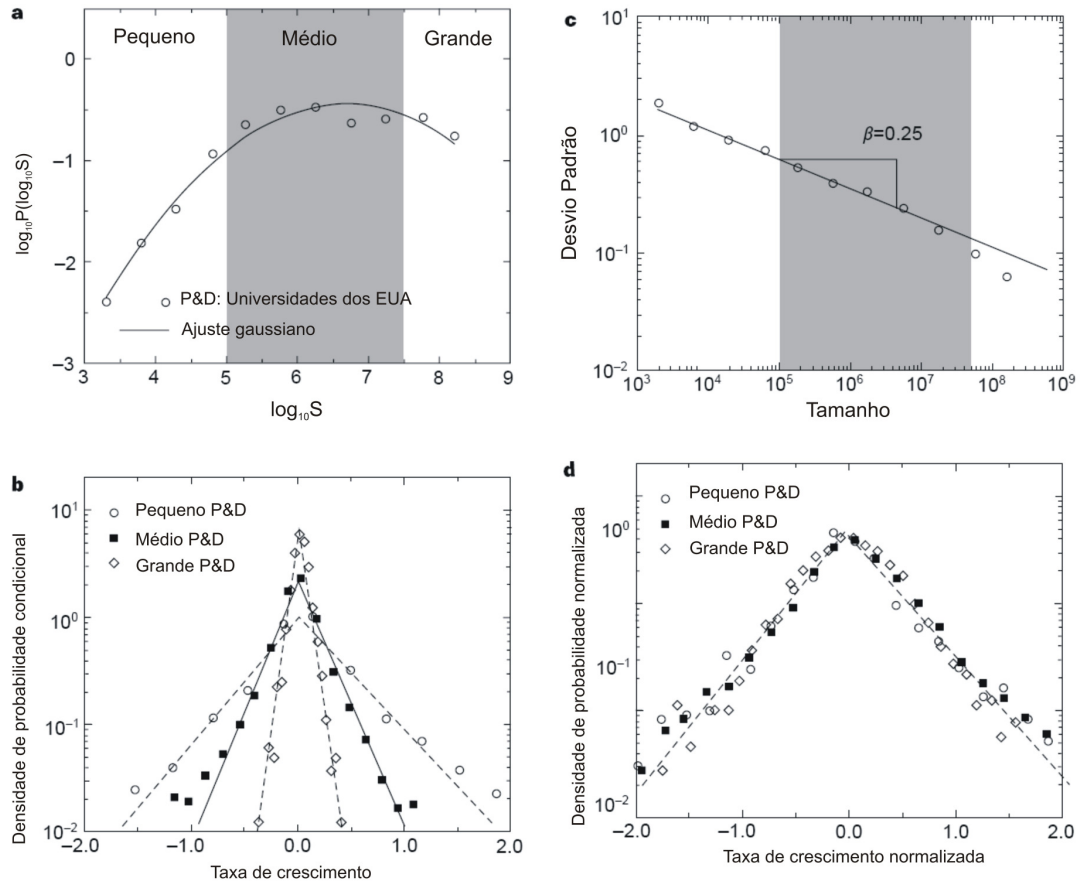


Figura 3.5: Figura retirada e adaptada da referência [23]. Em a), um gráfico do logaritmo da distribuição de probabilidades do logaritmo dos gastos com Pesquisa e Desenvolvimento (P&D) nos EUA é apresentado. A linha sólida representa um ajuste gaussiano. Em b), as densidades de probabilidades para três grupos distintos usando como critério os valores de seus gastos com P&D, a mesma forma funcional é encontrada. Como em b) as larguras são diferentes, esperamos que o desvio padrão decresça com o aumento do tamanho dos gastos, verificado em c). Realizada a normalização, em d), as distribuições de probabilidade dos três grupos colapsam em uma única curva.

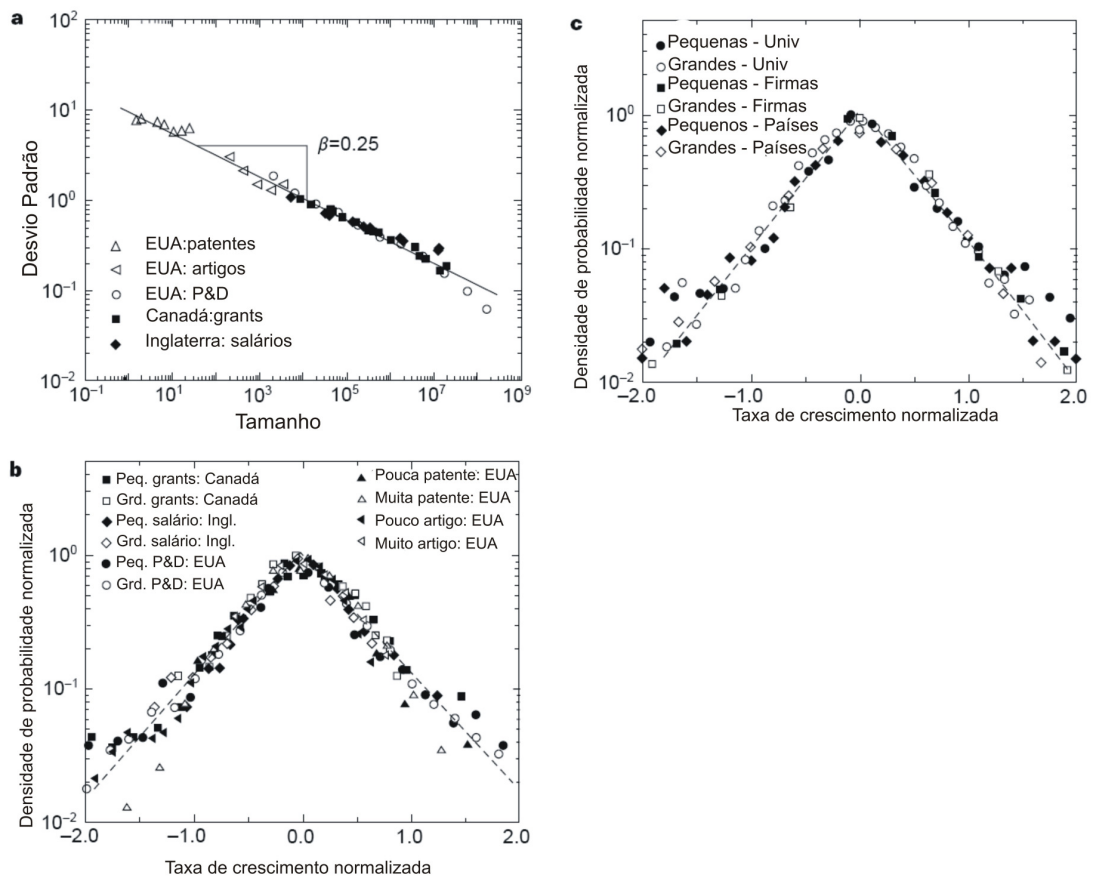


Figura 3.6: Figura retirada e adaptada da referência [23]. Em a), vemos o desvio padrão das taxas de crescimento para diferentes setores de diferentes países. A linha sólida indica uma lei de potência com expoente igual a $-0,25$. A idéia de universalidade é reforçada em b), onde a distribuição da densidade de probabilidade normalizada para os diferentes setores de diferentes tamanhos e de diferentes países colapsam em uma única curva. Em c), o colapso se dá na distribuição da densidade de probabilidade para universidades, para firmas e para países.

-0,25, figura 3.6a. A mesma forma funcional também é encontrada para a distribuição de probabilidade condicional $p(r|S)$, ilustrado na figura 3.6b.

Comparou-se a dinâmica de crescimento das universidades com o crescimento de firmas [17] e com o crescimento dos países [21] e notou-se que as distribuições da densidade de probabilidade normalizadas apresentam as mesmas características, representadas na figura 3.6c.

3.4 Produção Científica

Assim como o nosso trabalho, inúmeros outros são direcionados à investigação da produção bibliográfica; é o caso, por exemplo, das referências [7, 8, 9, 10, 11, 12].

Na revisão que segue, foram considerados três tipos diferentes de dados, um referente à publicação anual de 247 países entre 1980-2001, outro referente ao número total de artigos publicados anualmente por 508 institutos da União Européia (UE) e 408 institutos acadêmicos dos Estados Unidos da América (EUA) em um período de 11 anos (1991-2001) e, por último, o número de publicações de 2330 autores belgas.

No caso dos países e institutos acadêmicos, os dados foram extraídos a partir da versão CD-ROM do *Science Citation Index* (SCI) com informações e dados bibliográficos publicados entre os anos 1980-2001, publicado pelo *Institute for Scientific Information* (ISI). No caso dos autores, utilizou-se também uma versão CD-ROM do SCI, referente aos artigos, notas e revisões. Publicados entre os anos 1980-2000 pelos autores belgas.

A contagem das publicações dos países foi feita de três formas distintas. Na primeira, caso o artigo fosse produzido por autores de diferentes países, seria atribuído a cada país uma parte fracionária, consistindo do número de autores do referido país dividido pelo número total de autores, motivo pela qual é chamada de *contagem fracionária*. Suponha o exemplo, retirado da referência [9], consistindo de um artigo publicado por quatro autores, dois norte-americanos, um holandês e um belga. Assim, seria atribuído metade do artigo para os Estados Unidos, um quarto para Holanda e um quarto para Bélgica.

Na segunda, chamada de *contagem inteira, tipo I*, no mesmo caso do exemplo anterior, seriam atribuídos dois artigos para os EUA, um para Holanda e um para Bélgica. Na terceira e última, *contagem inteira, tipo II*, atribui-se um artigo para cada um dos países. É importante notar que nas contagens inteiras não há conservação do número de artigos, entretanto, os resultados são análogos para os três tipos de contagem.

Um total de 123 países foram excluídos do estudo por apresentarem produção muito baixa, ou mesmo nula, no intervalo considerado (1980-2001), restando somente 124 países. A quantidade de países assemelha-se com a de instituições que investigamos no capítulo seguinte deste trabalho, próximo de uma centena.

Inicialmente, têm-se as análises das publicações dos países. A figura 3.7 mostra o histograma do logaritmo do número de publicações dos 124 países, nos 22 anos considerados, utilizando a contagem fracionária. A distribuição exibe um comportamento bi-modal, sugerindo que os países podem ser divididos em duas classes, agrupados pelo tamanho de suas publicações. Análises com os outros tipos de contagens apresentaram resultados similares.

Dividiu-se os países em 10 grupos, de acordo com o magnitude do número de publicações S , e mais uma vez o desvio padrão, $\sigma(S)$, decai como uma lei de potência,

$$\sigma(S) \sim S^{-\beta}, \quad (3.15)$$

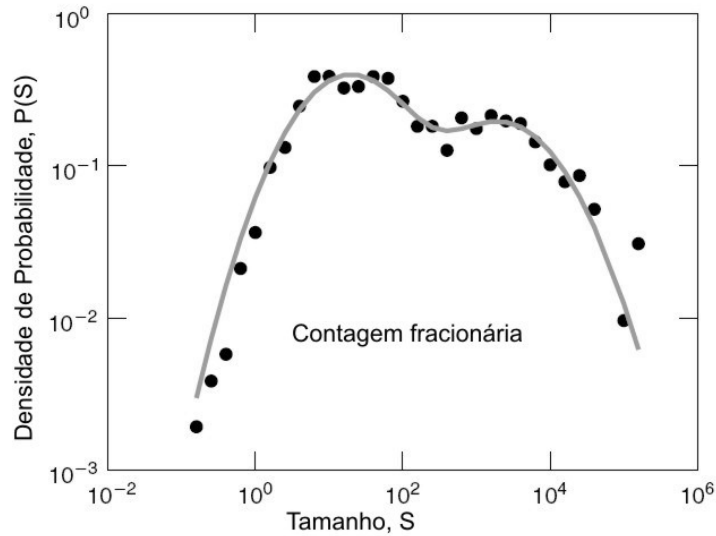


Figura 3.7: Figura retirada e adaptada da referência [9]. Como primeiro resultado tem-se uma distribuição de probabilidades bi-modal para o tamanho das publicações dos países, segundo a contagem fracionária. Os resultados obtidos com as contagens inteiras são análogos.

com $\beta \simeq 0,32$, veja a figura 3.8a.

Para verificar se as distribuições das taxas de crescimento apresentam a mesma forma funcional, os países foram divididos em três grupos e para cada grupo constrói-se um histograma normalizado das distribuições de probabilidades condicionais. Outra vez colapsam em uma única curva, com a forma de tenda, mostrado na figura 3.8b.

Em seguida, foi feita uma análise das publicações das instituições. A figura 3.9 mostra o histograma do logaritmo do número de artigos publicados pelas instituições dos EUA e da UE. Percebe-se, no caso das instituições européias, um comportamento uni-modal, em contraste com o bi-modal das instituições norte-americanas.

Também nota-se que o desvio padrão, $\sigma(S)$, decai segundo uma lei de potência com o número de publicações, S , como em (3.15), agora com $\beta = 0,39$ para instituições européias e $\beta = 0,30$ para as norte americanas, figura 3.10.

E, finalmente, foram analisadas as publicações dos autores belgas. A distribuição do logaritmo do número de artigos publicados apresenta um comportamento gaussiano uni-modal, figura 3.11. A distribuição do número de artigos publicados é aproximadamente lognormal, fazendo, mais uma vez, referência ao modelo de Gibrat.

A figura 3.12a representa o desvio padrão, $\sigma(S)$, que decai como uma lei de potência, equação (3.15), em função do número de artigos publicados (S), com expoente $\beta = 0,28$. Na figura 3.12b, tem-se a distribuição de probabilidades normalizada, o que mostra que para os diferentes autores a forma da distribuição se mantém a mesma.

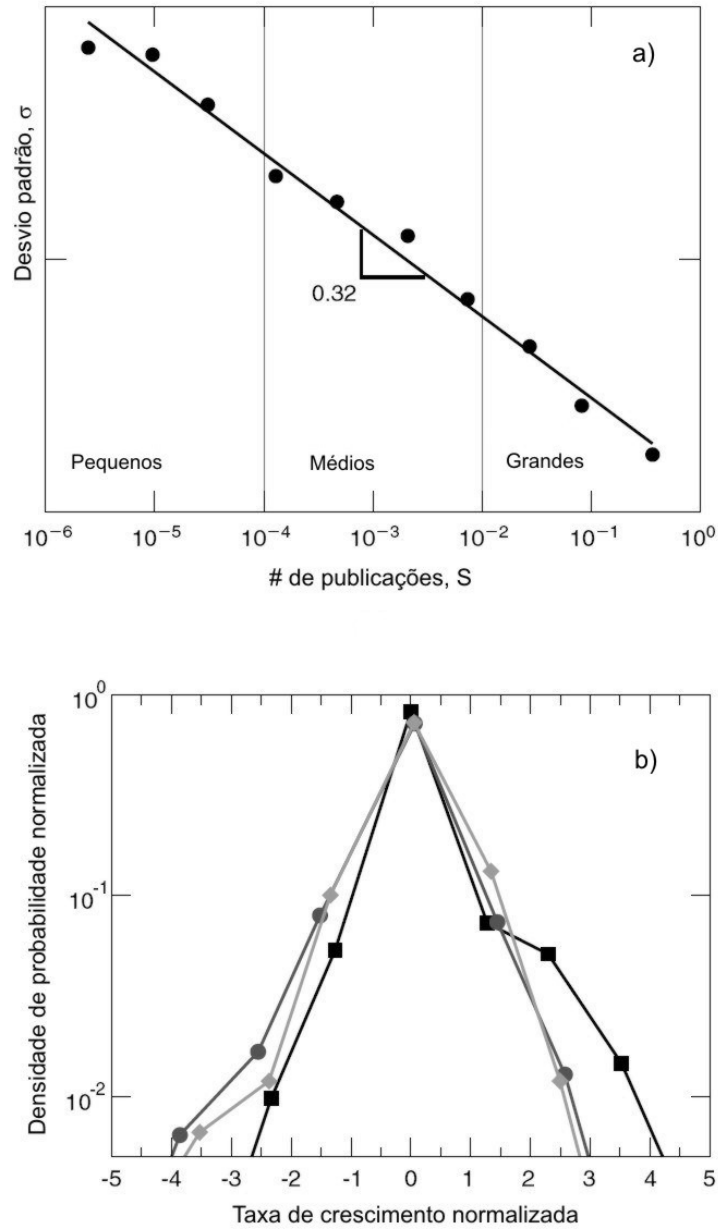


Figura 3.8: Figura retirada e adaptada da referência [9]. Vemos, em a), o decaimento do desvio padrão, σ , com o aumento do número de publicações, S . Em b), as distribuições de probabilidades normalizadas para os três diferentes grupos. Note o colapso dos dados em uma única curva com a forma de tenda (*tent-shape*).

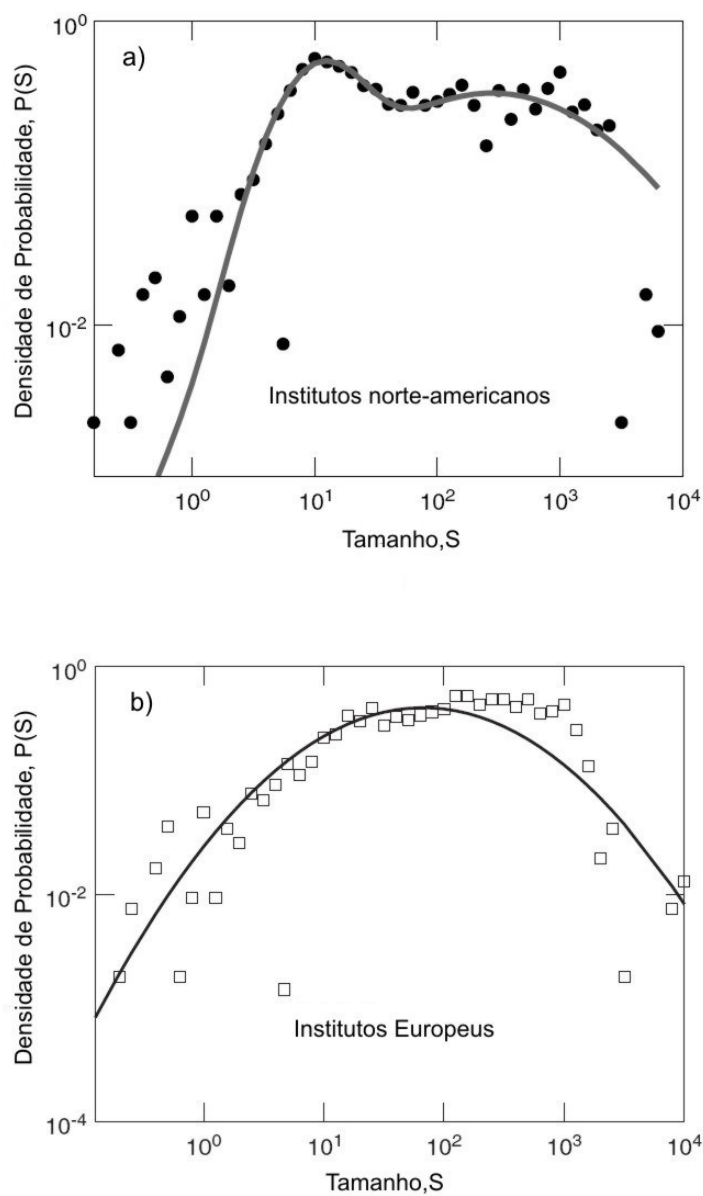


Figura 3.9: Figura retirada e adaptada da referência [9]. Histograma do logaritmo do tamanho da publicação de: a) 408 institutos dos Estados Unidos e b) 508 institutos da União Européia, medidos com o sistema de contagem fracional, para o período de 11 anos (1991-2001).

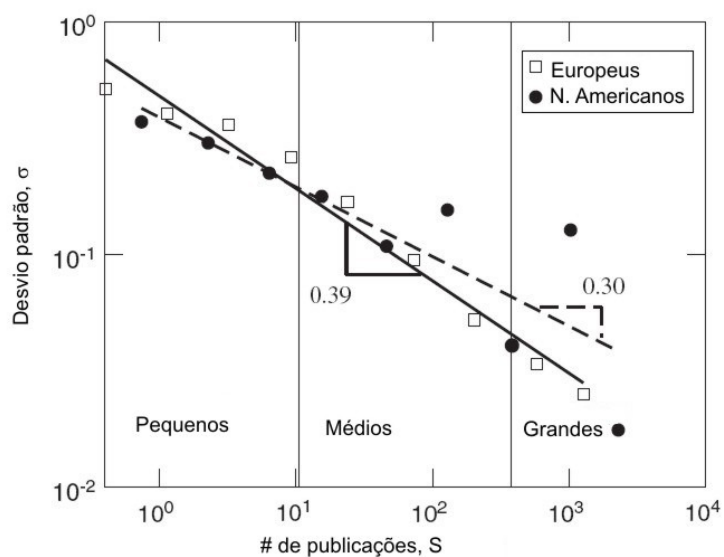


Figura 3.10: Figura retirada e adaptada da referência [9]. As instituições européias e norte-americanas foram divididas em 10 grupos de acordo com o tamanho de suas publicações, S . Percebemos então, que o desvio padrão, $\sigma(S)$, decai segundo uma lei de potência com expoente $-0,3$ para os institutos norte-americano (círculos) e $0,39$ para os institutos europeus (quadrados).

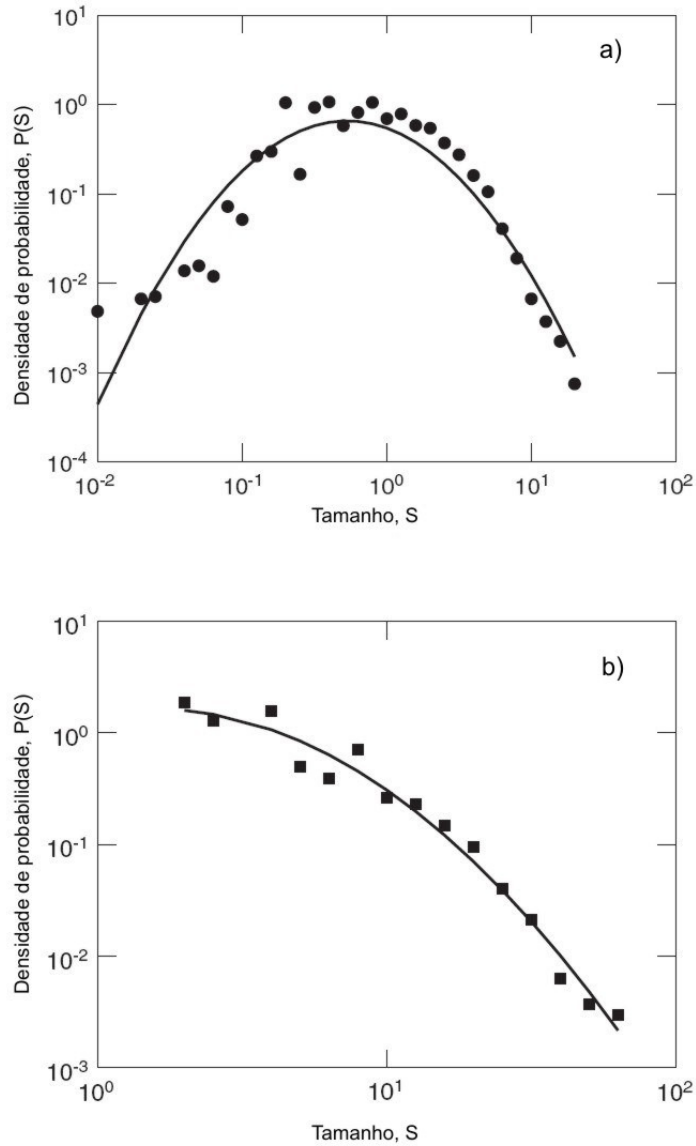


Figura 3.11: Figura retirada e adaptada da referência [9]. Histograma do logaritmo do número de publicações de 2330 autores belgas ao longo de 22 anos (1980-2001) para o sistema de contagem fracionária a) e inteira b). As linhas cheias são ajustes gaussianos, assim como os previstos por Gibrat, para sistemas em que as taxas de crescimento são independentes e não correlacionadas com o tamanho do sistema, com distribuição de probabilidades lognormal.

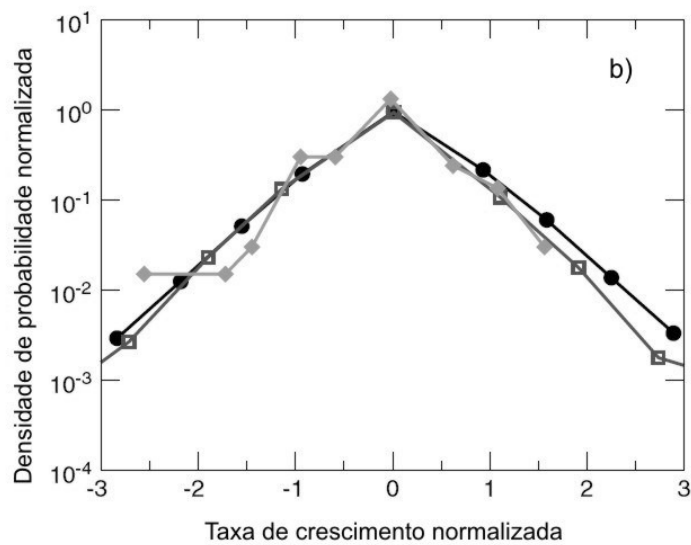
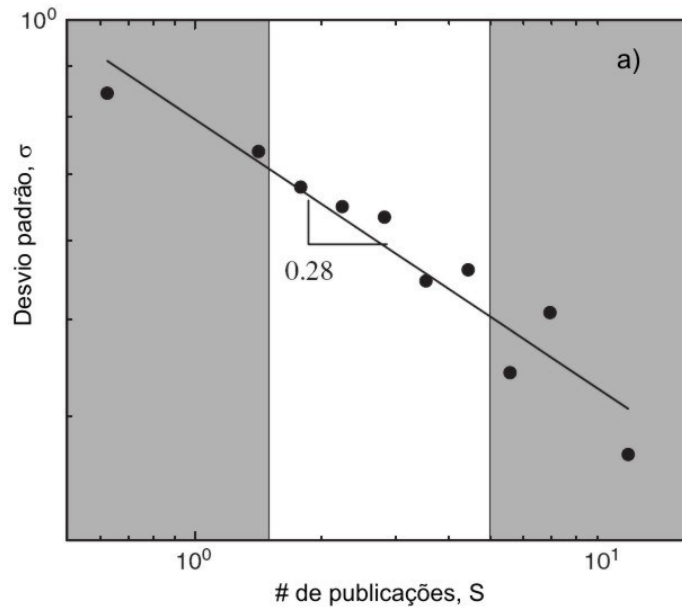


Figura 3.12: Figura retirada e adaptada da referência [9]. Utilizando o sistema de contagem fracionário para as publicações dos autores belgas, em a), foram divididos em 10 grupos de acordo com o tamanho de suas publicações. Foi encontrado que o desvio padrão, σ , decai segundo uma lei de potência de expoente $-0,28$. Já em b), a partir da distribuição de probabilidades normalizadas para as taxas de crescimento, vê-se o colapso de três diferentes setores.

Capítulo 4

Produção Científica do Brasil: Centros de Pesquisa

Este capítulo contém uma análise da produção científica brasileira. De um total de mais de duas mil e trezentas instituições de ensino superior [77] e institutos de pesquisa, foram encontrados cento e cinco com número de publicações consideráveis para nosso estudo, no sentido de apresentar regularidade de produtividade em algum intervalo dentre os anos investigados.

Obtivemos a produtividade anual dos centros de pesquisa brasileiros, ao longo de vários anos, desde sua primeira publicação até o ano de 2004, indicadas pelo *ISI* [1]. De uma maneira geral, seguimos a linha dos trabalhos apresentados nos capítulos anteriores e, em particular, esta investigação se assemelha à feita por *Matia et al* [9] em 2005, que trata da produção científica de institutos de pesquisa dos Estados Unidos da América e da União Européia, dos países e de autores belgas.

4.1 Introdução

Os dados foram extraídos da *internet* a partir do *site Web of Science* (WOS) [1], sob responsabilidade do *Institute for Scientific Information* (ISI). Para tentar evitar que diferentes formas de digitação e abreviação ou mesmo eventuais erros de ortografia influenciassem em nossa contagem, procedemos da forma mais detalhada: verificamos todos os artigos pesquisados. Tal procedimento favoreceu a obtenção confiável dos resultados. Ao invés de nos restringirmos ao nome da instituição, foi possível pesquisar pela sigla, pela cidade e pelo estado.

Vale ressaltar que contamos as publicações da seguinte forma: considere uma publicação com quatro autores, um da universidade X, outro da universidade Y e outros dois da universidade Z. Então, contou-se um artigo publicado por X, um para Y e também um para Z. Deixaremos mais detalhes sobre os métodos da pesquisa e a interface do *site* para o Apêndice C.

4.2 Métodos Utilizados

As análises feitas neste trabalho dividem-se em quatro partes. Na primeira, ajustamos o crescimento de todos os institutos por uma função exponencial. O primeiro passo foi

dispor os institutos em ordem decrescente da sua produtividade total ao longo dos anos estudados. Em uma pré-análise, verificamos que o tipo de crescimento predominante entre os maiores institutos era o exponencial. Assim, para o ajuste das restantes, deu-se preferência para a função exponencial e, então, foram feitos os gráficos para todas as instituições juntamente com a função exponencial que mais se aproximou dos pontos obtidos, utilizando o método dos mínimos quadrados.

Foi investigada também a distribuição do tamanho das instituições, representadas pelo número de publicações. Dando continuidade, calculamos as taxas logarítmicas de crescimento para anos consecutivos para todas as instituições, sendo possível determinar a forma da distribuição das mesmas. E, finalmente, investigamos o comportamento do desvio padrão com relação a média da magnitude de publicações de cada instituição, procedimentos esses listados e exemplificados no capítulo anterior.

De forma geral, as análises seguiram dois caminhos. No primeiro, focamos o tipo de crescimento, o exponencial, e optamos por investigá-lo para todas as instituições. No segundo, estudamos as propriedades exemplificadas no capítulo 3, oriundas em trabalhos envolvendo dinâmicas de crescimento.

4.3 Comportamento Exponencial

Os resultados encontrados evidenciam um crescimento das publicações brasileiras com o passar dos anos. Isso se justifica, entre vários fatores, pelo aumento do número de pesquisadores, pelo aumento do número de periódicos, pelo surgimento dos cursos de pós-graduação e também pelo incentivo das instituições científicas. Tal situação sugere a possibilidade de um crescimento exponencial, como nos exemplos vistos no capítulo 1.

Embora o crescimento das publicações científicas de cada instituição seja esperado, há casos em que isso não é tão claro. Trata-se, em geral, de instituições com um baixo número de publicações. Nesses casos, as flutuações se tornam bem mais notáveis e então pode ser menos óbvio a tendência de aumento de produtividade. Isso se deve ao fato de que em pequenas produtividades a ocorrência de poucos eventos pode representar uma fração significativa da produtividade. Por outro lado, flutuações relativamente menores ocorrem para aquelas instituições em que a produtividade anual passa de algumas centenas. De uma maneira geral, verificamos a tendência geral de que nas instituições com pequena produtividade, ou mesmo no início da produtividade das grandes instituições, existam grandes flutuações, que se tornam menos evidentes à medida que cresce o número de artigos publicados.

Alguns exemplos foram selecionados para ilustrar as flutuações, veja a figura 4.1. Nos gráficos aqui apresentados, utilizamos o ano 1 (um) para o primeiro ano em que a instituição tem um artigo registrado pelo WOS. Dessa forma, a medida que o ano aumenta, ele se aproxima e finaliza em 2004, o ponto mais à direita.

Motivados pela tendência de crescimento discutida, optamos por ajustá-la por funções exponenciais para cada instituição, embora alguns não apresentaram resultado satisfatório devido às flutuações. Os gráficos de todas as instituições, em ordem decrescente de sua produtividade total, encontram-se no Apêndice D.

Notamos que, em poucos casos, se optássemos pelo ajuste do tipo logístico, este também se daria de forma aparentemente satisfatória. Ocorre que algumas instituições parecem ter diminuído o ritmo de crescimento. Entretanto, tal desvio pode ser um mero

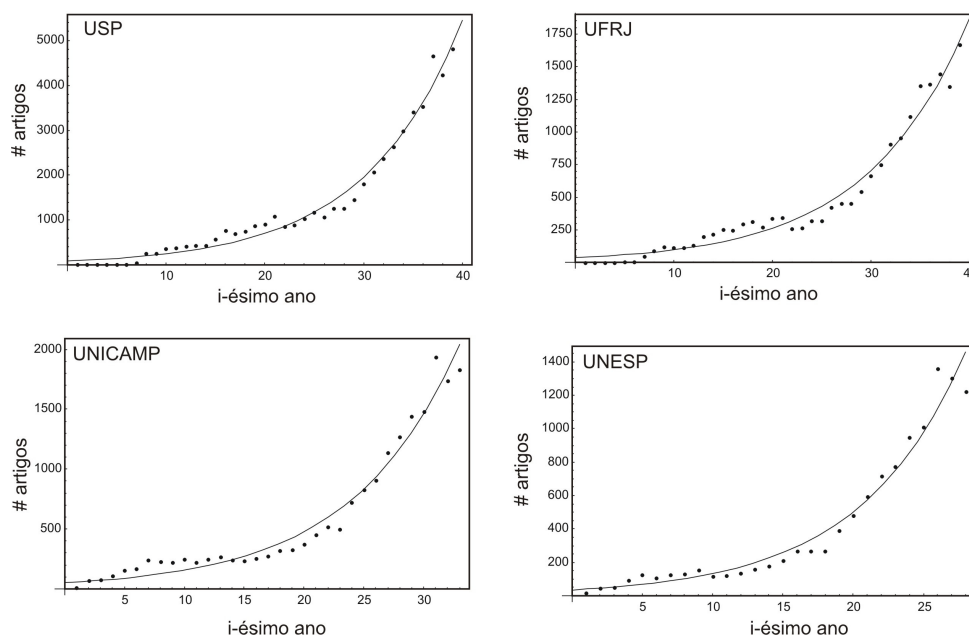


Figura 4.1: Ajuste exponencial para as quatro instituições científicas brasileiras mais produtivas (em número de artigos).

fruto de flutuações. A figura 4.2 ilustra instituições com ajuste logístico e exponencial para que possamos comparar.

Encontramos os resultados que passamos a descrever. Feito o ajuste exponencial, com uma função do tipo

$$s(t) = A \exp(Bt) , \quad (4.1)$$

aqui consideramos gráficos dos coeficientes A 's e B 's, para todas as instituições pesquisadas, em função da ordem da produtividade total. De fato, o coeficiente A decai com a diminuição da produtividade e o coeficiente B tem uma tendência de crescimento com a diminuição da produtividade de cada instituição de pesquisa brasileira e comporta-se como constante, em média, ao considerarmos somente as mais produtivas. Notamos novamente, com a diminuição da produtividade, um aumento nas flutuações dos coeficientes. As figuras 4.3 e 4.4 mostram tais situações. Por sua vez, deve ser notado que a figura 4.4, quando restritas aos casos de baixa flutuação (os 52 maiores centros de pesquisa), indicam que os coeficientes B 's tendem a uma constante, isto é, $B = 0,12$ (com duas casas decimais de precisão).

Se olharmos para a produtividade (em artigos no *ISI*) do Brasil, identificamos uma boa concordância com uma evolução do tipo exponencial (ver a figura 4.5). Também nesse caso, $B = 0,12$ (duas casas decimais de precisão). Esse fato mostra o que era de ser esperar, o comportamento médio das partes está em concordância direta com o comportamento do todo. Além disso, mostra de maneira quantitativa a tendência de crescimento da produtividade científica brasileira.

Uma visão mais detalhada da dinâmica do crescimento científico do Brasil deve levar em conta não apenas a tendência exponencial como um todo ou para cada centro de pesquisa, mas sim o comportamento ano a ano da produtividade de cada instituição científica. É justamente a esse tipo de análise que dedicamos o restante deste capítulo.

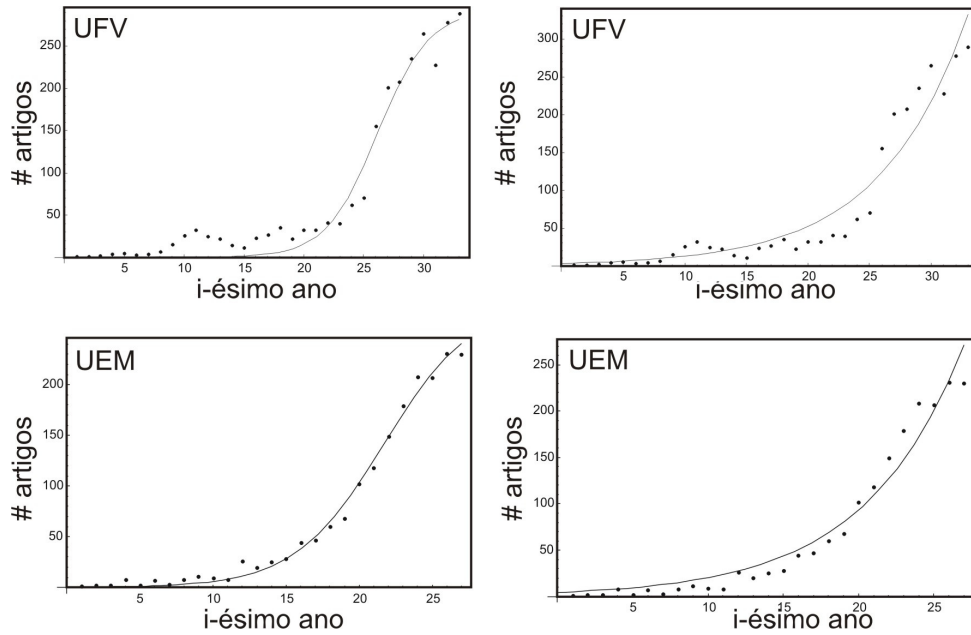


Figura 4.2: Embora optamos pelo ajuste exponencial, a produção científica de algumas instituições podem ser bem ajustadas por uma função logística. À esquerda o ajuste é feito com uma função logística e à direita com uma função exponencial, para as mesmas instituições.

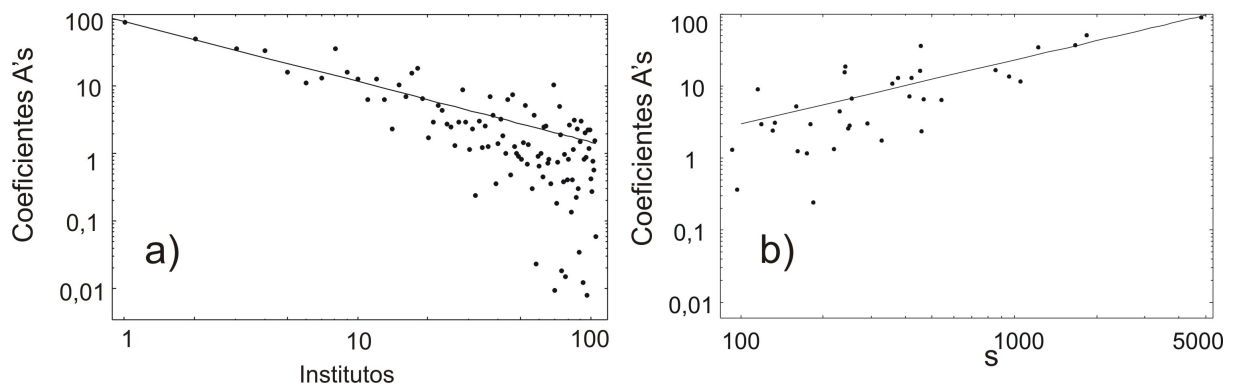


Figura 4.3: Em a), vemos como os coeficientes A 's apresentam um decaimento conforme diminui a produtividade dos institutos (classificação em ordem decrescente) e, em b), fica evidenciado o crescimento com o aumento da produtividade, aqui para as 36 instituições mais produtivas. A linha sólida à esquerda corresponde a lei de potência: $A(n) = 92,94 n^{(-0,90)}$, em que n é um número que representa a classificação do instituto, à direita indica também uma lei de potência, $A(s) = 0,50 s^{0,89}$, com s representando a produtividade no ano de 2004.

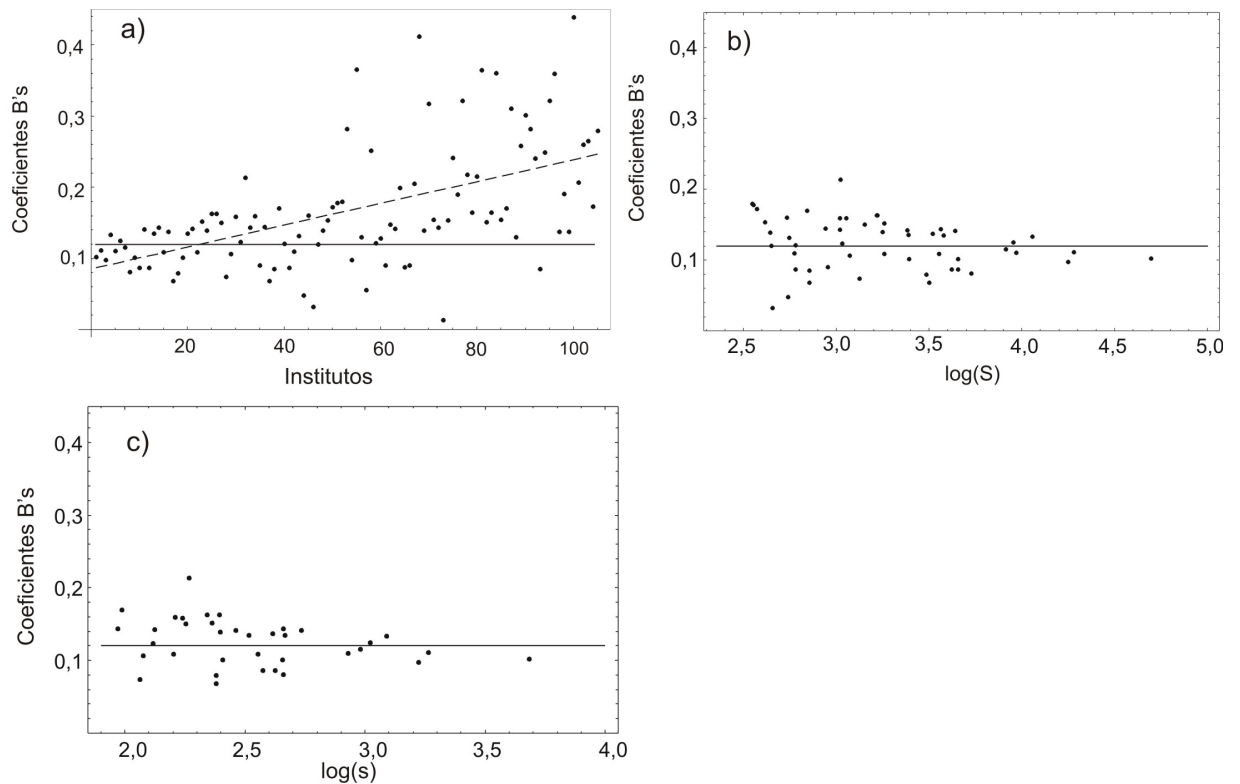


Figura 4.4: Em a), notamos que os coeficientes B 's apresentam uma leve tendência de crescimento, quando consideramos todas as instituições de pesquisa. Em contraste, quando nos restringimos às instituições mais produtivas, é evidenciada uma constância, em média, dos B 's. Para as instituições mais produtivas, em b), temos os coeficientes B 's em função do logaritmo da produtividade total de cada instituto, S . E finalmente, em c), os coeficientes em função do logaritmo da produtividade no ano de 2004, $\log(s)$, para as instituições mais produtivas. A linha tracejada corresponde a $B(n) = 0,085 + 0,001n$, e as linhas sólidas marcam 0,12, o valor médio dos coeficientes B 's

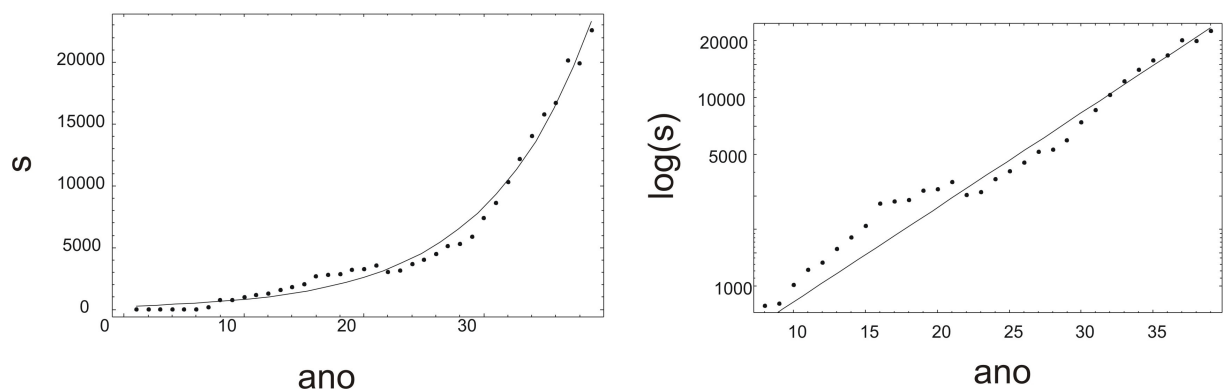


Figura 4.5: Evolução da produtividade científica do Brasil (em artigos) em gráfico usual, em a), e em gráfico mono-log, em b). Os pontos representam a produtividade brasileira em cada ano até 2004. As linhas contínuas correspondem a função $s(t) = 258,63 \exp(0,12t)$.

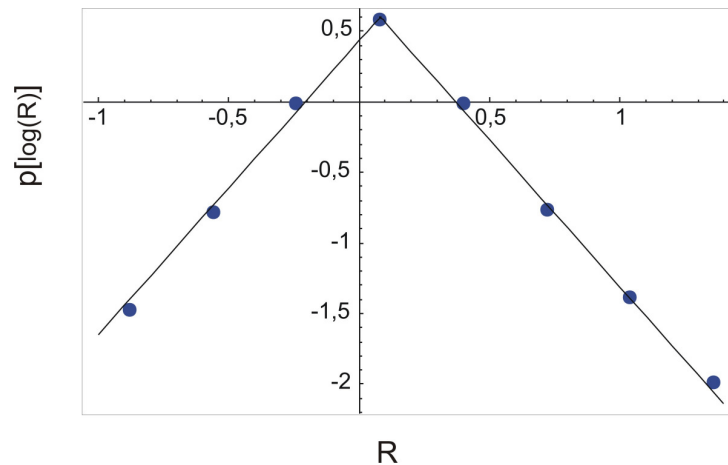


Figura 4.6: A distribuição de probabilidade da taxa de crescimento segue um padrão exponencial, como nos exemplos discutidos no capítulo anterior. Percebemos também que seu valor médio é positivo e, conseqüentemente, a tendência é que, em média, as instituições cresçam ao passar dos anos.

4.4 Distribuição das Taxas de Crescimento

A seguir, o caminho a ser trilhado se assemelha ao traçado por *Mathia et al* (2005) na investigação da produtividade de instituições norte-americanas e européias. Faremos uma comparação entre os resultados a fim de identificar o quão as dinâmicas de crescimento são semelhantes.

Analisaremos agora as taxas logarítmicas de crescimento (ou simplesmente taxas de crescimento) anuais, $R(t)$, para cada uma das instituições brasileiras relacionadas. Como já vimos, o procedimento consiste em construir uma nova série para cada instituição, cujos termos são definidos como o logaritmo da razão entre o número de artigos publicados, $s(t + 1)$, no ano $t + 1$, e o número de artigos publicados no ano anterior, t , exatamente como em (2.7),

$$R(t) = \log \left[\frac{s(t + 1)}{s(t)} \right] .$$

Em instituições com pouca produtividade anual, não é raro surgirem indeterminações nos $R(t)$, seja por divisão por zero, ou por tomarmos o logaritmo de zero. Tais indeterminações são desconsideradas em nossa análise.

Com todas as séries formadas pelas taxas de crescimento das instituições de pesquisa brasileiras, encontramos a distribuição de probabilidades para $R(t)$, indicada na figura 4.6.

O próximo passo foi, ao invés de levar em conta todas as instituições, considerar o grupo das 20 primeiras instituições e também o grupo das 20 últimas, desprezando as cinco menores, pois apresentavam grandes flutuações. Analogamente ao que foi feito anteriormente, fizemos simultaneamente o gráfico das distribuições de probabilidades desses dois grupos, observando que a forma da distribuição é mantida. Aqui nosso objetivo é mostrar que o tamanho das instituições determina a largura da distribuição (ver figura 4.7). Cabe lembrar que este resultado é observado também em outros sistemas, como na dinâmica de crescimento do PIB dos países [21, 22], no crescimento de firmas [17], dos

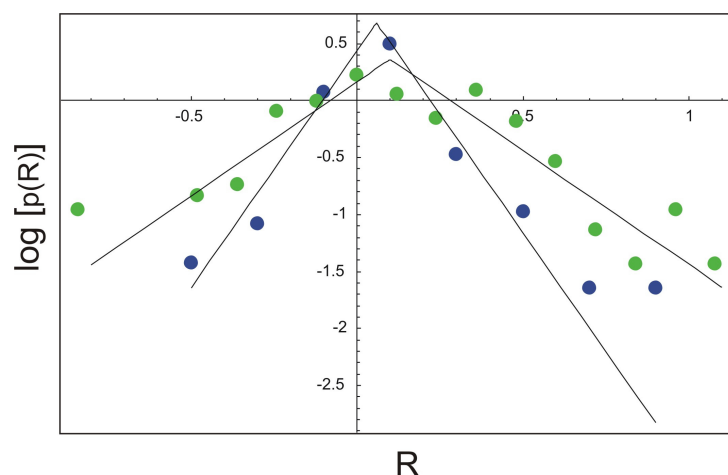


Figura 4.7: Em azul, a distribuição de probabilidades das taxas de crescimento das vinte maiores (primeiras) instituições (em número total de artigos publicados). Em verde, temos a distribuição das taxas de crescimento também de vinte instituições, desde a octagésima primeira até a centésima, segundo a mesma classificação. A forma da distribuição é mantida embora as larguras sejam diferentes, o que sugere um desvio padrão maior nas menores instituições. As curvas contínuas correspondem a ajustes usando a equação (2.8).

investimentos em pesquisa e desenvolvimento de universidades [23] (ver figura 3.5), crescimento do fator de impacto de periódicos [12] (ver figura 3.2) e também com o crescimento das instituições estudadas por *Mathia et al* (2005) [9] (ver figura 3.8).

Embora a distribuição de probabilidades tenha a mesma forma de tenda, elas apresentam diferentes larguras. Como já mencionado, essa característica se deve ao fato de que o desvio padrão tende a ser maior no caso das menores instituições, justificado pelo aumento das flutuações. Assim, quanto maior a universidade, menores as flutuações e, portanto, mais estreita a distribuição de probabilidades. Para reforçar as discussões a respeito do comportamento universal em dinâmicas de crescimento, constatamos que, ao normalizar as distribuições de probabilidades, para as grandes e pequenas instituições brasileiras, elas colapsam em uma única curva, demonstrando a semelhança na forma de crescimento (veja figura 4.8). Esse resultado está novamente em concordância com a investigação realizada por *Mathia et al* (2005)[9] levando em conta tanto outras instituições como diversos outros países e autores. Essa concordância também se dá com os resultados das referências do parágrafo anterior.

4.5 Distribuição do Tamanho das Instituições

Em relação à distribuição do tamanho das instituições, fizemos um histograma para representá-la. Em virtude dos resultados apresentados em capítulos anteriores, gostaríamos de verificar se ela segue uma distribuição lognormal e, por essa razão, o gráfico de $p(\log s)$ deveria ser comparado com o de uma parábola. Devido aos relativamente poucos dados, o nosso ajuste poderia apresentar-se com muitas flutuações em relação a curva de ajuste, porém com uma escolha conveniente da largura das janelas do histograma esse tipo de discrepância pode ser minimizado. O histograma com a distribuição dos tamanhos (número de publicações) das instituições brasileiras para o ano de 2004, usando uma largura de

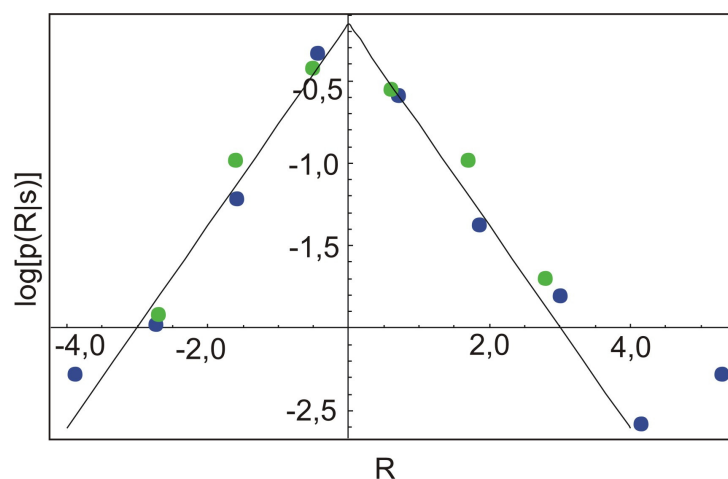


Figura 4.8: Com a normalização (uso de $r = (R - \mu)/\sigma$), a distribuição das taxas de crescimento colapsam numa única curva. A curva contínua representa a equação (2.8) com desvio padrão unitário, $\sigma = 1$ e média nula, $\mu = 0$.

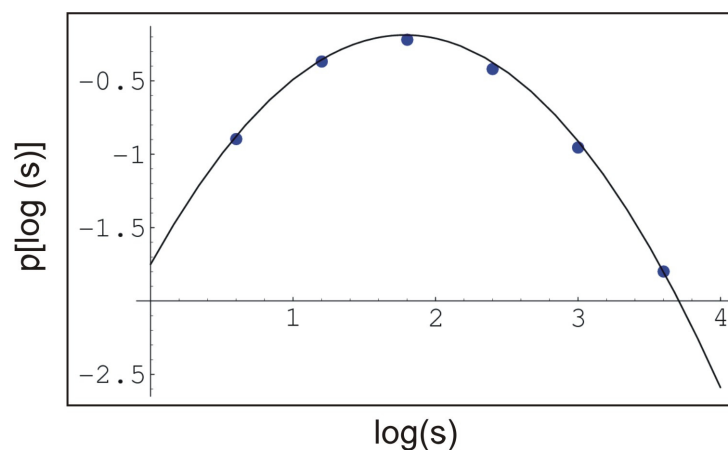


Figura 4.9: Percebemos que a distribuição dos tamanhos das instituições de pesquisa brasileiras é aproximadamente lognormal, visto que é bem ajustada por uma parábola. A janela utilizada, na variável $\log[s]$, foi de 0,6.

janela conveniente, é mostrado na figura 4.9.

Como podemos notar, nosso resultado, baseados no ano de 2004, indica que a distribuição do tamanho (quantidade de artigos publicados) das instituições de pesquisa brasileiras é lognormal. Resultados similares foram verificados para outros anos. Isso mostra que esse resultado é similar ao das instituições de pesquisa da União Européia (ver figura 3.9b), em contraste com as do Estados Unidos da América (ver figura 3.9a).

4.6 Desvio Padrão das Taxas de Crescimento

Por fim, analisamos os desvios padrões das taxas de crescimento da produtividade das instituições de pesquisa brasileiras, a partir das quais deveríamos encontrar uma diminuição desta grandeza à medida que aumenta a produtividade. Para tal, podemos proceder de formas diferentes. Podemos dividir as instituições em vários grupos, de acordo

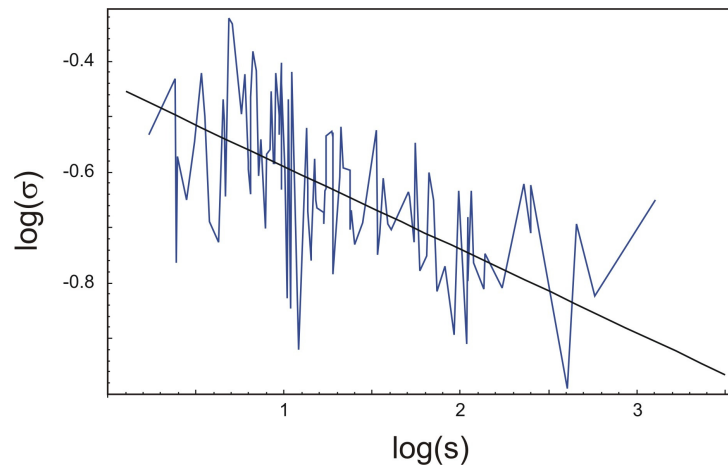


Figura 4.10: Há uma tendência de decréscimo do desvio padrão à medida que aumenta a produtividade dos institutos. Esse decréscimo segue uma lei de potência, com expoente $-0,15$.

com o número de publicações, e para cada grupo calcular o desvio padrão das respectivas taxas de crescimento. Outra forma é, para cada grupo, calcular a média dos desvios padrões das taxas de crescimento de cada instituição. No trabalho de *Mathia et al* (2005), as instituições foram divididas em 10 grupos, e para cada grupo foi determinado o desvio padrão das taxas de crescimento e então fez-se o gráfico do desvio padrão, σ , em função do tamanho (número de publicações), S . O resultado encontrado foi a dependência do tipo lei de potência desses desvios com o tamanho da instituição. Essa forma é especialmente útil, pois diminui as flutuações como pode ser visto a partir da figura 3.8a.

Utilizamos uma forma alternativa, calculamos o desvio padrão para cada universidade e fizemos o gráfico em função do tamanho destas. Embora as flutuações sejam maiores, fica claro o decaimento do desvio padrão com o aumento do tamanho. A figura 4.10 mostra nossos resultados. Optamos por este método devido ao pequeno número de dados e a produtividade não se distribuir de forma uniforme. Poderíamos ter um grupo com um número de instituições acima da média e outro com um número bem abaixo e, assim, as flutuações continuariam presentes em alguns desses grupos.

Como resultado, percebemos que há um decaimento, e que quando aproximado pela lei de potência

$$\sigma(s) \propto s^{-\beta}, \quad (4.2)$$

conduz a $\beta = 0,15$. Note que este valor dá uma medida quantitativa das flutuações presentes no sistema em estudo.

Capítulo 5

Conclusão

Neste trabalho, apresentamos uma breve discussão, via exemplos, a respeito de séries temporais são um tópico comumente abordado em física estatística. Ressalta-se que aplicações desse tipo de estudo, que contém muitos dados, são enormemente facilitadas devido ao desenvolvimento da informática, situação que não ocorreria há alguns anos. Aqui, empregando séries temporais, colocamos em destaque o estudo de dinâmicas de crescimento, representado, entre outras características, pelo tamanho de firmas, tamanho do PIB dos países, tamanho do investimento em pesquisa e, obviamente, pela quantidade de publicações científicas de institutos científicos, que é o tema central desta dissertação.

Em primeiro lugar, foram expostas algumas distribuições comumente encontradas nesses tipos de estudos. Junto às distribuições, há alguns exemplos ilustrando diversas áreas em que tais situações foram obtidas. Essa apresentação (capítulo 1), poderia ser omitida para os que têm formação básica em aspectos gerais de probabilidade, porém, os exemplos escolhidos ressaltam o aspecto interdisciplinar das leis de crescimento e distribuições de probabilidade por nós empregadas.

Já no segundo capítulo, enfatizamos alguns conceitos de séries e flutuações e como eles podem ser utilizados na análise estatística de certos sistemas particulares. Nesse contexto, incluem-se análises encontradas em trabalhos de física estatística; mais precisamente, expomos o essencial de séries temporais empregado nesta dissertação, assim como, ilustrações do uso de análise de flutuações.

Com esses estudos preliminares, torna-se clara uma semelhança entre crescimentos, independentemente do sistema em questão. Tais similaridades sugerem que organizações distintas, como firmas, países e universidades, são candidatas a serem governadas por mecanismos de crescimento comuns ou bem parecidos. Nesse sentido de mecanismos de crescimento comuns, apresentamos o modelo de Gibrat, um modelo inicialmente proposto para crescimento de firmas que se mostrou importante em nossa discussão da dinâmica da produção científica brasileira por reproduzir a distribuição do tamanho das publicações, além da dinâmica de outros sistemas. Essas características comuns, portanto, nos conduzem a um cenário em que estão presentes dois conceitos muito empregados no contexto de sistemas complexos: o de invariância de escala e o de universalidade.

Vimos no capítulo 3 a invariância de escala na distribuição exponencial (Laplace) das taxas de crescimento do fator de impacto de periódicos científicos, nas taxas de crescimento de circulação de revistas e jornais, nas taxas de crescimento de gastos em pesquisa em universidades, nas taxas de crescimento de firmas, nas taxas de crescimento do PIB e, por fim, nas taxas de crescimento da produção científica de países, institutos e autores.

A mesma invariância é notada para a distribuição do tipo lei de potência para o desvio padrão das taxas de crescimento. Dito de outra forma, esse capítulo mostra, de maneira detalhada, como a maioria das técnicas empregadas no capítulo central desta dissertação são implementadas, com ênfase no exemplo da produção científica norte-americana e européia.

Enfim, chegamos ao estudo da dinâmica de crescimento dos institutos de pesquisa brasileiros, o material discutido no capítulo 4, em que, investigamos quatro aspectos:

1 - Cada instituição tem sua produtividade, em artigos científicos, bem representada por uma função exponencial?

2 - A comparação entre a quantidade de artigos científicos publicados por centros de pesquisa: distribuição das produtividades.

3 - A distribuição das taxas logarítmicas de crescimento da produção científica para as diversas instituições de pesquisa.

4 - A comparação entre os desvios padrões das taxas de crescimento logarítmica para os vários centros de pesquisa.

Verificamos que a maioria das instituições de pesquisa tem sua produtividade, em função do tempo, bem representada por um crescimento exponencial. Exceção deve ser feita, em geral, para as instituições com baixa produtividade, já que grandes flutuações relativas na produtividade podem ocorrer para centros com poucos artigos. Assim, vários ajustes por exponenciais podem não manifestar bem o comportamento científico da instituição. De uma maneira geral, quando representamos o comportamento exponencial por $s(t) = A \exp(Bt)$ para cada instituição, vimos que os A 's são, em média, crescentes com a produtividade e os B 's aproximadamente constantes. Esse último fato, sugere um comportamento universal de crescimento para as instituições científicas brasileiras, sendo inteiramente compatível com o comportamento científico do Brasil como um todo, pois o valor encontrado para o ajuste da produtividade nacional, ano a ano, foi de 0,12, que é o valor médio dos B 's encontrados para as instituições de pesquisa brasileiras.

Quando fizemos um histograma da quantidade de centros de pesquisa em função da produtividade, vimos que essa distribuição é bem representada por uma distribuição log-normal. Tal fato indica a possibilidade de conectar esse sistema a um processo multiplicativo que pode ser conectado diretamente ao modelo de Gibrat, que tem sido usado na modelagem do tamanho de firmas. O modelo supõe que as suas diversas partes (aqui representados pelos centros de pesquisas) evoluem segundo um processo multiplicativo, porém sem interação entre elas. Este último ingrediente do modelo indica que, efetivamente, não há interferência entre os distintos centros de pesquisa. Dito de outra forma, parece não haver um plano diretor (comportamento coletivo) que faz com que as instituições tenham um comportamento coerente (interdependente). O comportamento mostra-se semelhante ao caso das instituições de pesquisa da União Européia e em contraste com a natureza bi-modal dos centros de pesquisa dos Estados Unidos da América.

As distribuições das taxas logarítmicas de crescimento da produção científica (em artigos) mostraram uma forma do tipo tenda, num gráfico mono-log, o que vem se mostrando bastante comum para vários sistemas estudados, como ficou claro a partir dos exemplos apresentados nos capítulos 2 e 3. Tal fato sugere um tipo de universalidade e não um padrão particular oriundo das instituições de pesquisa brasileiras. Aqui podemos nos referir ao modelo estocástico proposto por Picoli Jr. por reproduzir essa característica.

Uma outra vertente relacionada à taxa logarítmica de crescimento das produções científicas, que é comumente investigada, refere-se aos desvios padrões relativos às partes

(aqui os centros de pesquisa). Seus valores, em função da produtividade, sugerem um comportamento do tipo lei de potência embora na nossa investigação contenha poucas partes (centros de pesquisa). Vale mais uma vez lembrar o modelo estocástico referenciado no parágrafo anterior, visto que esta característica é também reproduzida. Por isso e motivado por outros sistemas com muitos elementos, que apresentam um comportamento do tipo lei de potência, fomos conduzidos a um expoente $-0,15$. Agora, cabe ressaltar que esse expoente é aproximadamente a metade daqueles encontrados no caso das instituições de pesquisa dos Estados Unidos da América e da União Européia.

Os comportamentos acima, munidos de todas as propriedades qualitativas e quantitativas, servem para caracterizar o nosso sistema (produtividade científica dos centros de pesquisa brasileiros) e assim, se desejarmos, podemos compará-lo com outros. Nesse ponto, alguém poderia questionar quanto a nossa base de dados, no sentido que ela não representa a totalidade da produção científica do mundo. De fato, não estão presentes livros, artigos científicos ausentes do ISI (maior parte da produção mundial) etc. Contudo, é a base que temos a maior facilidade de acessar. Além disso, é de se esperar, numa primeira estimativa, que os dados do ISI sejam uma fração aproximadamente fixa da produção verdadeira.

A seguir, passemos a discutir as perspectivas de avanços e novas direções relativas ao nosso trabalho. Para tal, é conveniente fazermos uma breve retrospectiva das vertentes de pesquisa do grupo no qual fiquei inserido neste mestrado. O grupo vem se dedicando a investigar temas em mecânica estatística e, em particular, conectadas a mecânica estatística generalizada, difusão anômala e mais recentemente ao estudo de séries temporais. Esta última com uma visão dominante na linha deste trabalho. Dessa forma, como perspectiva natural para investigações futuras, cabem duas vertentes: uma é a de estudar outros aspectos da produção científica (direção que não foi objetivo do presente trabalho) e a outra é a de usar as técnicas deste trabalho para investigar outros sistemas. Apesar de não ter propostas claras do que fazer especificamente nessas duas direções de pesquisa, elas ditam as nossas ambições imediatas para dar prosseguimento a este trabalho.

Apêndice A

Teorema Central do Limite

O teorema central do limite[78] afirma que a variável aleatória w , definida por

$$w = \frac{1}{\sqrt{N\sigma^2}} \left\{ \sum_{j=1}^N x_j - N\mu \right\} , \quad (\text{A.1})$$

possui a distribuição de probabilidades gaussianas

$$\frac{1}{\sqrt{2\pi}} e^{-w^2/2} , \quad (\text{A.2})$$

no limite $N \rightarrow \infty$. Além disso, para que o teorema central do limite seja válido, basta que as variáveis aleatórias x_j sejam independentes, existir a média μ dessas variáveis, assim como sua variância σ^2 .

Para obter (A.2), calcularemos primeiramente o valor médio de e^{ikw} , denotado por $\langle e^{ikw} \rangle$. Ademais, por simplicidade de notação, usaremos $\mu = 0$. Nessa direção, observemos inicialmente que a independência das variáveis aleatórias x_j conduz a

$$\begin{aligned} \langle e^{ikw} \rangle &= \left\langle \exp \left[ik \frac{\sum_{j=1}^N x_j}{\sqrt{N\sigma^2}} \right] \right\rangle \\ &= \left\langle \exp \left(ik \frac{x_1}{\sqrt{N\sigma^2}} \right) \right\rangle \dots \left\langle \exp \left(ik \frac{x_N}{\sqrt{N\sigma^2}} \right) \right\rangle \\ &= \left\langle \exp \left(ik \frac{x}{\sqrt{N\sigma^2}} \right) \right\rangle^N , \end{aligned} \quad (\text{A.3})$$

pois considera-se que todas variáveis x_j têm distribuições de probabilidades idênticas. Por sua vez,

$$\begin{aligned} \left\langle \exp \left(ik \frac{x}{\sqrt{N\sigma^2}} \right) \right\rangle &= \int_{-\infty}^{\infty} \exp \left(ik \frac{x}{\sqrt{N\sigma^2}} \right) p(x) dx \\ &= g \left(\frac{k}{\sqrt{N\sigma^2}} \right) , \end{aligned} \quad (\text{A.4})$$

em que

$$g(s) = \int_{-\infty}^{\infty} e^{isx} p(x) dx . \quad (\text{A.5})$$

Usualmente, $g(s)$, que é a transformada de Fourier de $p(x)$, é conhecida como a função característica de $p(x)$.

A partir de $g(s)$, obtemos, por exemplo, que

$$\begin{aligned} g(0) &= \int_{-\infty}^{\infty} p(x)dx = 1 , \\ \left. \frac{dg}{ds} \right|_{s=0} &= \int_{-\infty}^{\infty} ixp(x)dx = i\langle x \rangle , \\ \left. \frac{d^2g}{ds^2} \right|_{s=0} &= \int_{-\infty}^{\infty} (ix)^2 p(x)dx = -\langle x^2 \rangle . \end{aligned} \quad (\text{A.6})$$

Daí, podemos calcular a variância em termos das derivadas de $g(s)$, pois

$$\sigma^2 = \langle (x - \langle x \rangle)^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2 . \quad (\text{A.7})$$

Das equações (A.6) e (A.7), segue que

$$\begin{aligned} g(s) &= g(0) + \left. \frac{dg}{ds} \right|_{s=0} s + \frac{1}{2} \left. \frac{d^2g}{ds^2} \right|_{s=0} s^2 + \dots \\ &= 1 + i\langle x \rangle s - \frac{1}{2} \langle x^2 \rangle s^2 + \dots \\ &= 1 - \frac{1}{2} \sigma^2 s^2 + \dots . \end{aligned} \quad (\text{A.8})$$

Aqui usamos $\mu = \langle x \rangle = 0$.

Esse resultado substituído na equação (A.4), quando usamos $s = k/\sqrt{N\sigma^2}$, é especialmente útil para $N \gg 1$ já que

$$\begin{aligned} \left\langle \exp \left(ik \frac{x - \mu}{\sqrt{N\sigma^2}} \right) \right\rangle &= \left(1 + -\frac{\sigma^2}{2} \frac{k^2}{N\sigma^2} + \dots \right) \\ &= 1 - \frac{k^2}{2N} + \dots \\ &\simeq 1 - \frac{1}{N} \frac{k^2}{2} . \end{aligned} \quad (\text{A.9})$$

Note que a parte desprezada nessa aproximação é proporcional a $N^{-3/2}$ e, portanto, desprezível em comparação com o resto da expressão para N muito grande.

Ao empregar o resultado (A.9), vê-se que a equação (A.3) pode ser aproximada por

$$\begin{aligned} \langle e^{ikw} \rangle &\simeq \left(1 - \frac{1}{N} \frac{k^2}{2} \right)^N \\ &\simeq e^{(-k^2/2)} , \end{aligned} \quad (\text{A.10})$$

pois $e^u = \lim_{N \rightarrow \infty} (1 + u/N)^n$.

Visto que a função característica $\langle e^{ikx} \rangle$ é a transformada de Fourier da distribuição de probabilidade da variável w , obtemos essa distribuição calculando a transformada inversa,

isto é,

$$\begin{aligned} P(w) &= \int_{-\infty}^{\infty} e^{-ikw} \langle e^{ikw} \rangle \frac{dk}{2\pi} \\ &\simeq \int_{-\infty}^{\infty} e^{-ikw} e^{-k^2/2} \frac{dk}{2\pi} \\ &= \frac{1}{\sqrt{2\pi}} e^{-w^2/2}, \end{aligned} \tag{A.11}$$

que é a equação (A.2).

Apêndice B

Distribuição Lognormal

Distribuições lognormais podem ser geradas por processos multiplicativos [79], e que Gibrat [63] chamou de efeito de proporcionalidade. Tais processos são caracterizados por uma sucessão de valores, em que cada valor consiste do seu anterior multiplicado por um número aleatório. Suponha uma grandeza z cujo valor inicial seja $z(0)$. Mas a cada passo, a grandeza pode crescer (ou decrescer), dependendo da variável aleatória \tilde{R} , assim, de maneira simplificada podemos supor

$$z(i+1) = \tilde{R}(i)z(i) , \quad (\text{B.1})$$

e daí (equação 2.3) concluir que $z(n) = \tilde{R}(n-1) \dots \tilde{R}(0)z(0)$. Tomando o logaritmo de $z(n)$, chegamos a

$$\ln z(n) = \sum_{j=0}^{n-1} \ln \tilde{R}(j) + \ln z(0) . \quad (\text{B.2})$$

Seja $x_{j+1} = \ln \tilde{R}(j)$, então podemos reescrever (B.2) como

$$\ln \left[\frac{z(n)}{z(0)} \right] = \sum_{j=1}^n x_j . \quad (\text{B.3})$$

Visando aproveitar o teorema central do limite, vamos considerar a variável

$$w = \frac{\sum_{j=1}^N x_j}{\sqrt{N\sigma^2}} \quad (\text{B.4})$$

Segundo este teorema, concluímos que a distribuição de probabilidades para w tende à equação (A.2),

$$P(w) = \frac{1}{\sqrt{2\pi}} e^{-w^2/2} . \quad (\text{B.5})$$

Aqui, por simplicidade estamos usando média nula ($\mu = \langle x \rangle = 0$).

Seja $z = z(N)/z(0)$, o que conduz, via as equações (B.3) e (B.4), a

$$w = \frac{\ln z}{\sqrt{N\sigma^2}} \quad (\text{B.6})$$

e, conseqüentemente,

$$dw = \frac{1}{\sqrt{N\sigma^2}} \frac{1}{z} dz . \quad (\text{B.7})$$

Assim, $\widehat{P}(z) dz = P(w) dw$ implica

$$\begin{aligned}\widehat{P}(z) &= P(w) \frac{dw}{dz} \\ &= \frac{1}{\sqrt{2\pi N\sigma^2}} \frac{1}{z} \exp\left\{-\frac{[\ln(z)]^2}{2N\sigma^2}\right\},\end{aligned}\tag{B.8}$$

que é então a distribuição de probabilidades lognormal (1.17), com média nula e N absorvido numa redefinição de σ ($N\sigma^2 \rightarrow \sigma^2$).

Apêndice C

Coleta dos Dados

Nossa busca de dados via a *internet* foi realizada por meio do *site* do *Web of Science*, que nos permite diferentes tipos de pesquisa. Utilizamos o tipo *GENERAL SEARCH* (pesquisa geral), onde poderíamos obter diversas informações a respeito do artigo.

Um dos campos para pesquisa era o *ADDRESS* (endereço). Nesse campo era concentrado informações a respeito do nome da instituição, sigla, endereço postal, cidade, estado e país, entre outros. Era onde entrávamos com as informações para a nossa pesquisa. Por exemplo, ao pesquisar o estado do Paraná, entramos com as siglas das universidades: “ufpr, uel, uem, etc”. Como surgiam resultados indesejáveis, optamos por aumentar a restrição, visto que verificaríamos todos os artigos. Assim, indicamos também as cidades: “curitiba, londrina, maringa, etc”, o estado: “parana” e o próprio nome das instituições nas línguas portuguesa e inglesa.: “Univ Fed Parana, Univ Est Londrina, Univ Est Maringa, Fed Univ Parana, State Univ Londrina, State Univ Maringa, etc”.

A forma de buscar os resultados, na realidade, faz uso de operadores chamados *booleanos*. Assim, quando optamos por buscar várias cidades, devemos usar o operador OR (ou). A busca por siglas deve ficar assim: “ufpr or uel or uem or ...”, ou seja, mostre os resultados de “ufpr” e também os resultados de “uel” e também “uem” e assim por diante. Já quando desejamos fazer a vinculação da sigla ao nome do país devemos usar o operador AND (e), e somente será mostrado os resultados em que as duas palavras são encontradas. Exemplo: Entrando com os valores “ufpr and brazil or uel and brazil or uem and brazil or ...”, teremos somente os resultados em que, no mesmo artigo, sejam encontradas a sigla da instituição UFPR e o país BRAZIL, também a sigla da instituição UEL e o país BRAZIL e também a sigla da instituição UEM e o país BRAZIL, mas é importante ressaltar que não há um campo específico para cada informação, a sigla, o nome da instituição, a cidade, o estado, o país o código postal e outras informações são dispostos a critério do autor no campo *ADDRESS* (endereço). Na figura C.1, temos a página principal do *site* de pesquisa empregado na nossa pesquisa. Quanto ao nosso processo detalhado de coleta de dados, ele durou aproximadamente 6 meses.

ISI Web of KnowledgeSM [Preview the New Version](#) Web of Science GO [HOME](#) [LOG OUT](#)

[Web of Science®](#) [Information for New Users](#) ✓ MY START PAGE

Select a search option:

Quick search: Example: tuberculosis vaccine*

Author Finder: Need help finding papers by an author? Use Author Finder. **NEW!**

Looking for a topic, author, or journal? Use General Search.

Open a previously saved search history.

Clear all search forms, your marked list, and your search history.

Select database(s) and timespan:

Citation Databases:

- Science Citation Index Expanded (SCI-EXPANDED)--1945-present
- Social Sciences Citation Index (SSCI)--1956-present
- Arts & Humanities Citation Index (A&HCI)--1975-present

Latest (updated August 25, 2007)

Year

From to (default is all years)

To remember these settings, first [sign in or register](#).

The Notices file was last updated 4/3/2007

Please give us your [feedback](#) on using the ISI Web of Knowledge.

[Acceptable Use Policy](#)
 Copyright © 2007 [The Thomson Corporation](#)

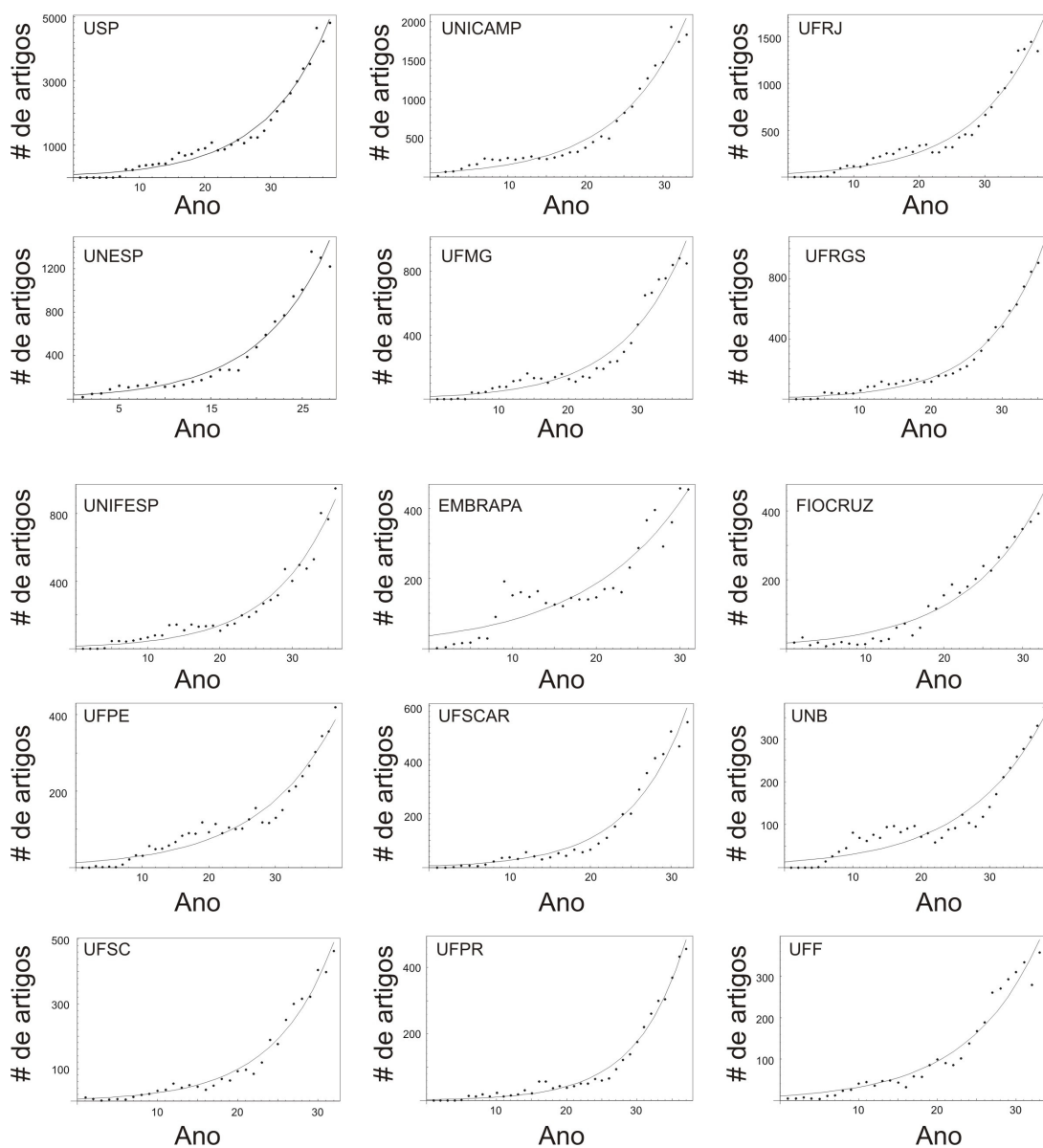
THOMSON

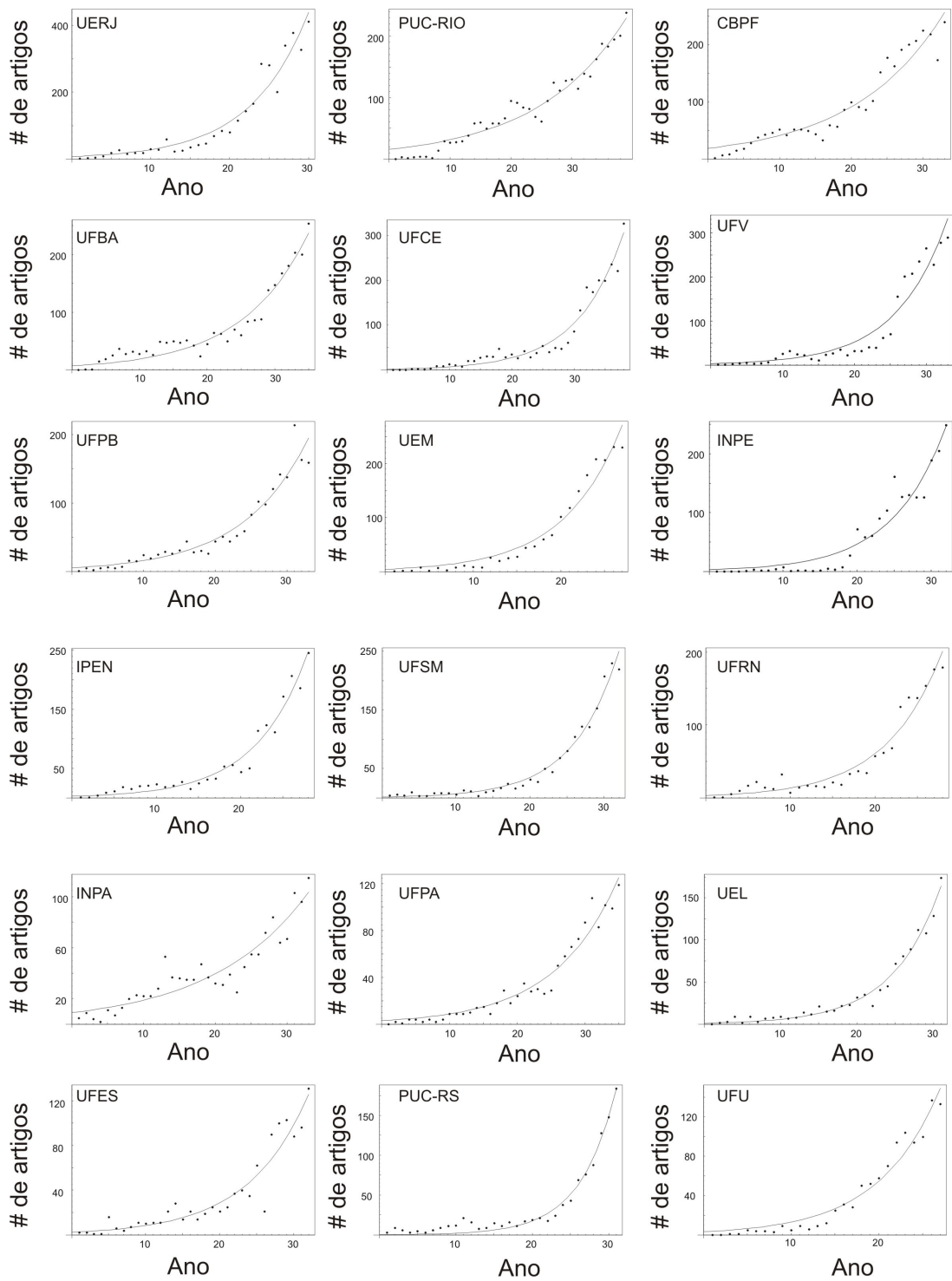
Figura C.1: Página inicial do Web Of Science usada durante a pesquisa.

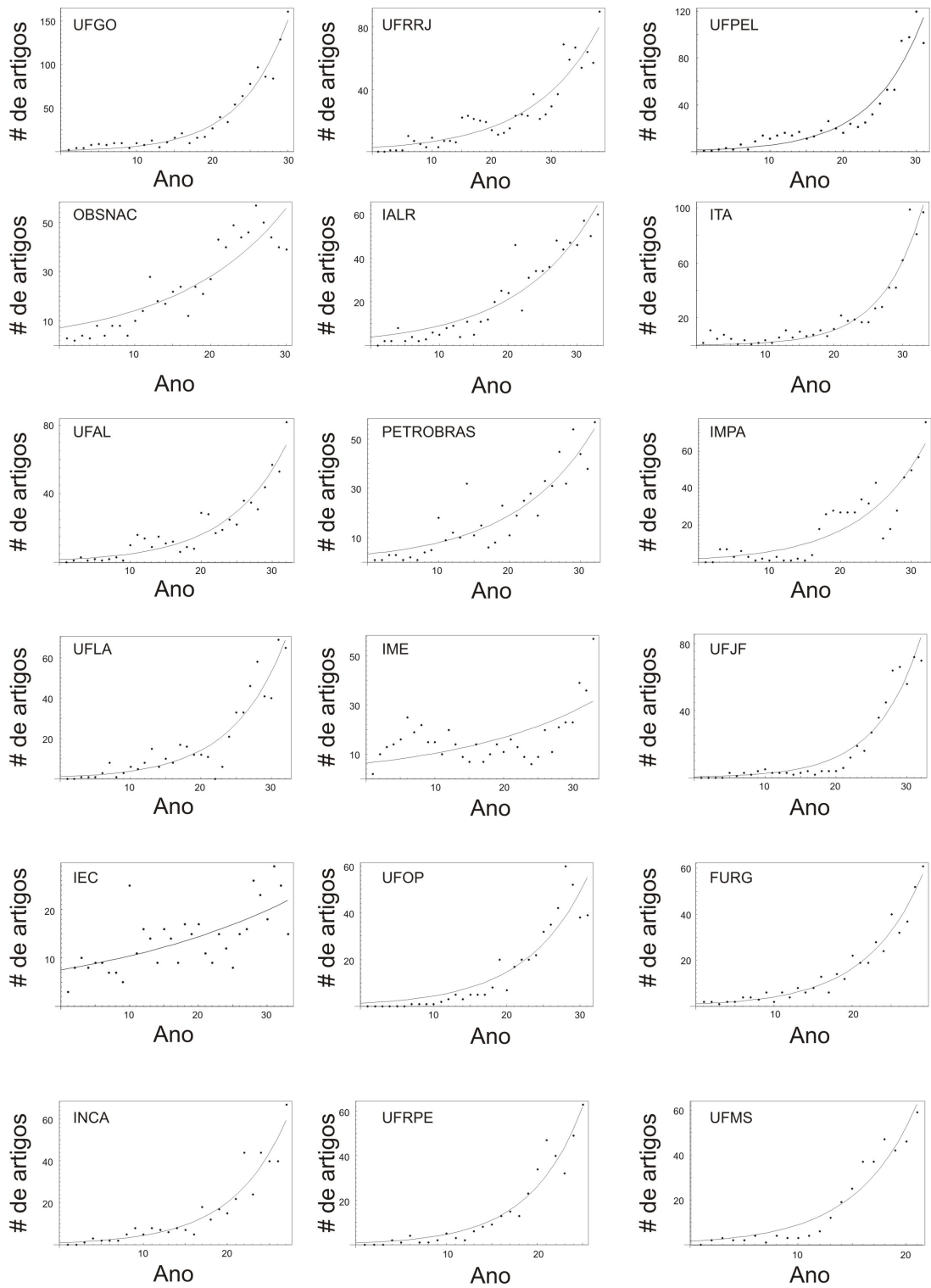
Apêndice D

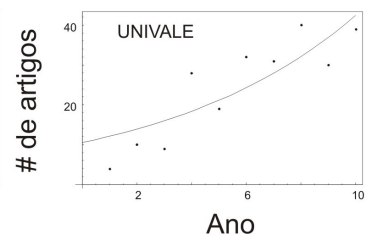
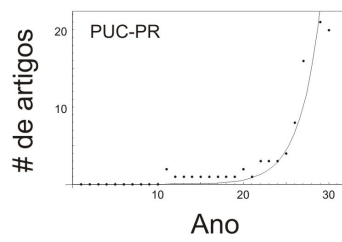
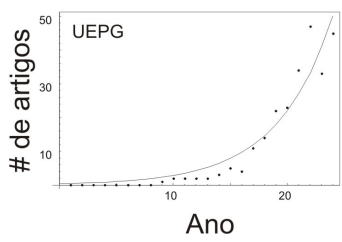
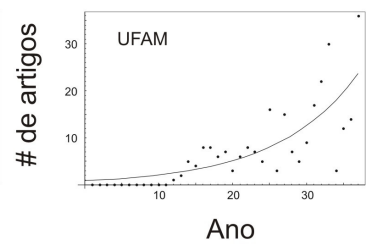
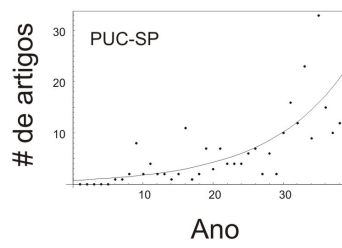
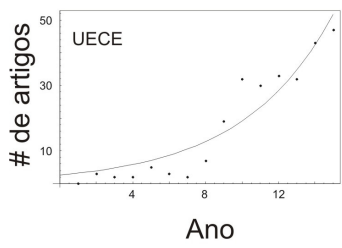
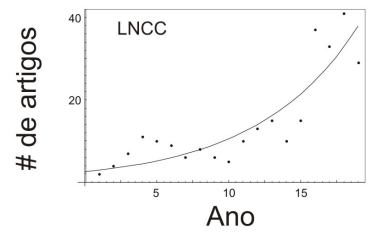
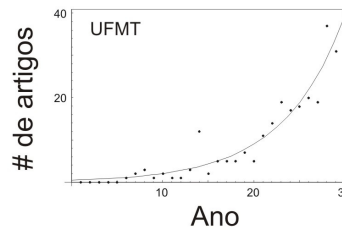
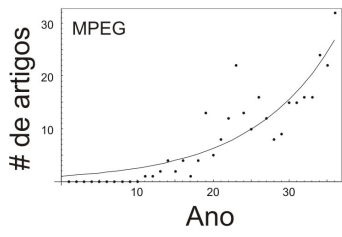
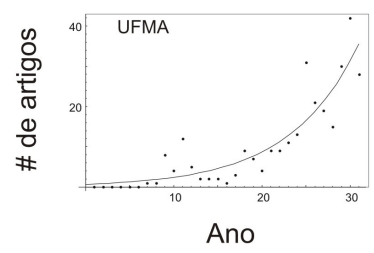
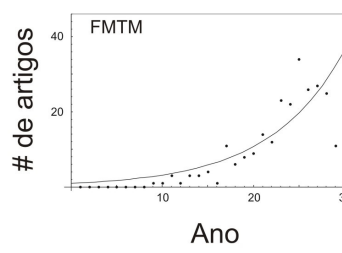
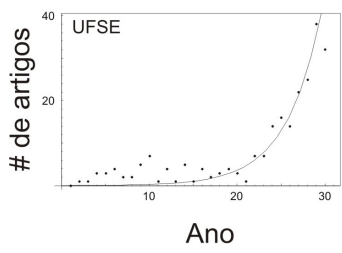
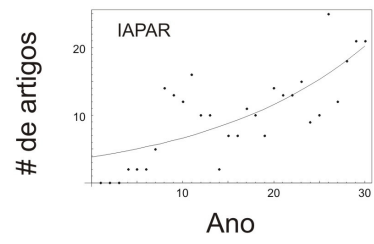
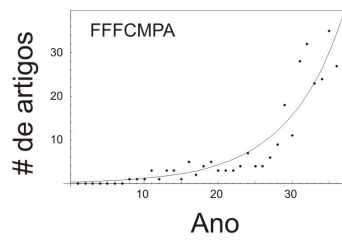
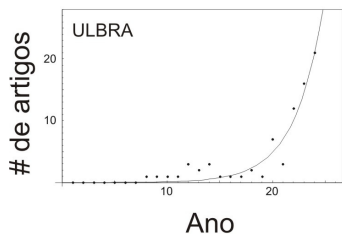
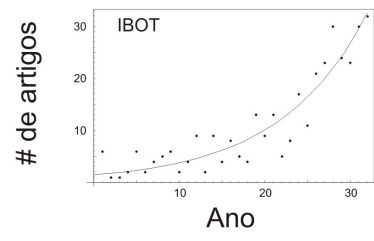
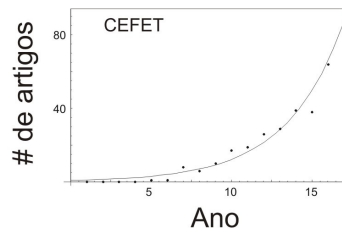
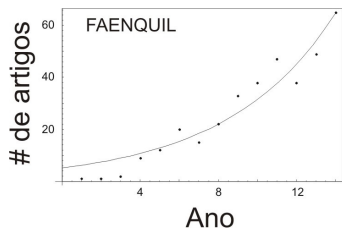
Gráficos

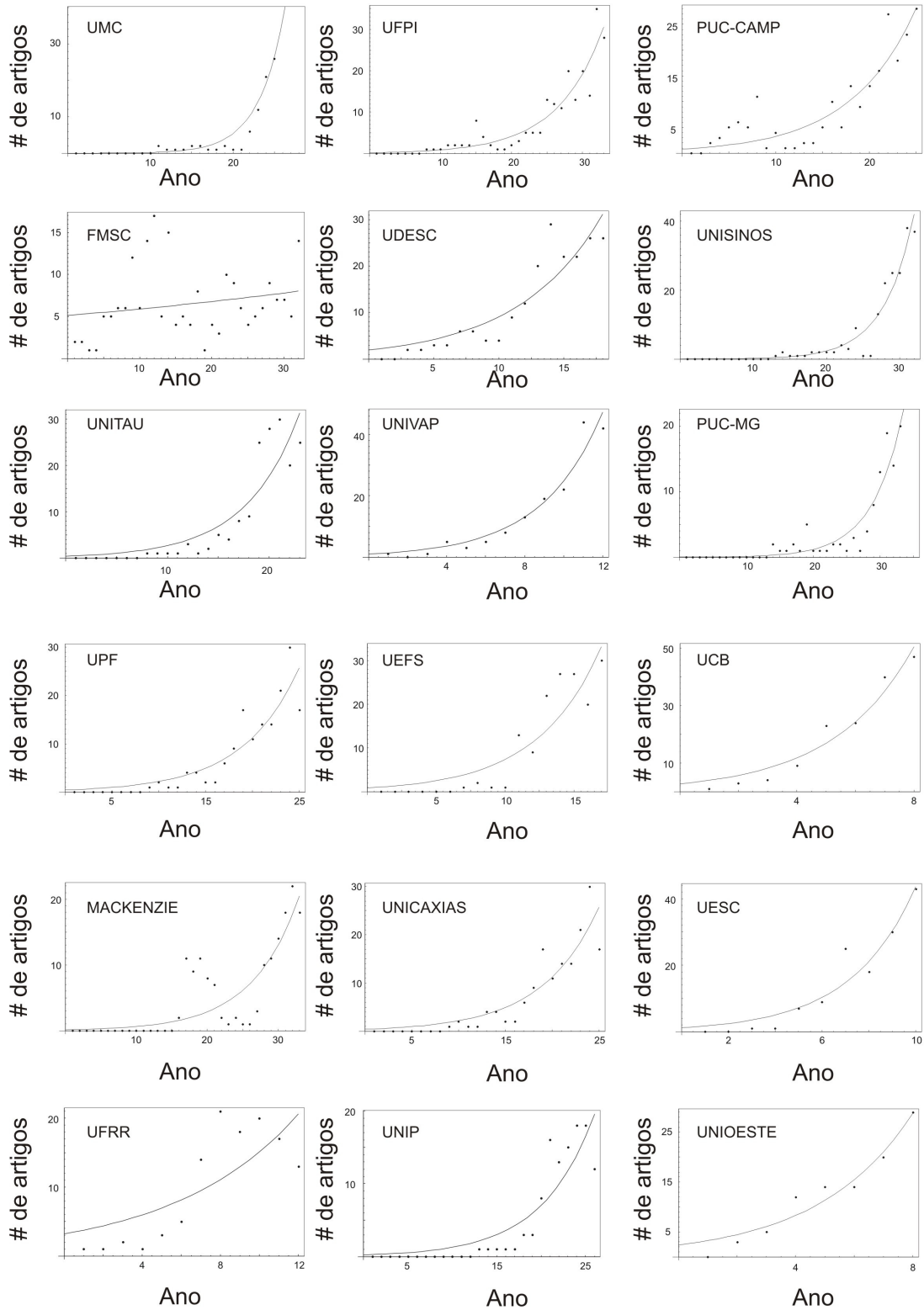
A seguir, apresentamos os gráficos da produtividade científica das 105 instituições brasileiras pesquisadas no *Web of Science*. Após os gráficos, relacionamos as siglas com os nomes das instituições.

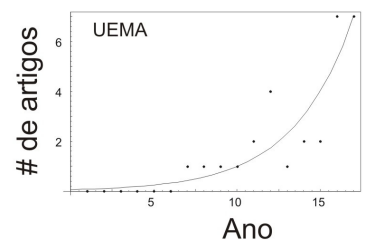
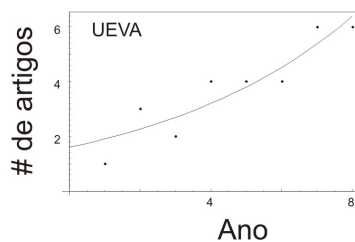
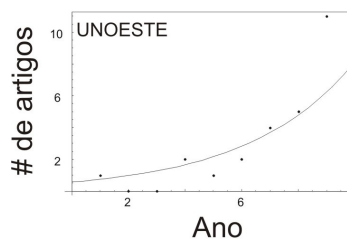
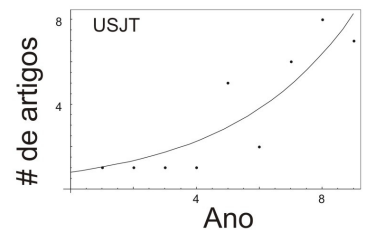
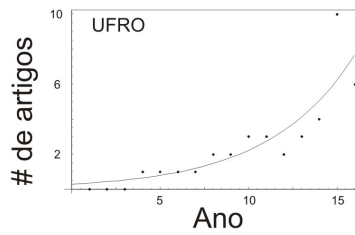
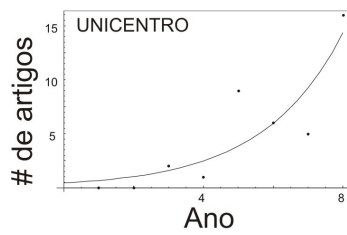
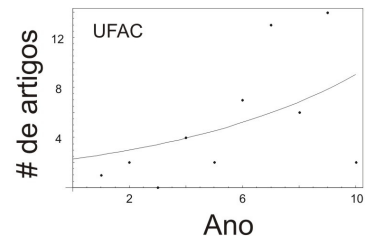
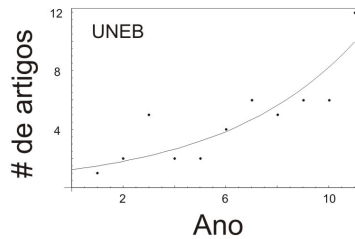
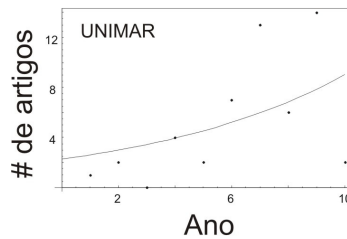
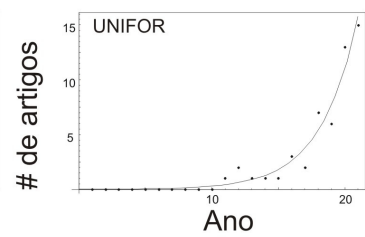
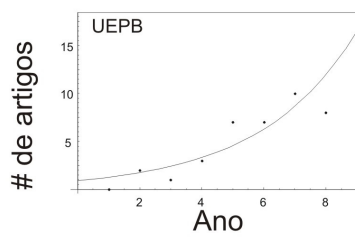
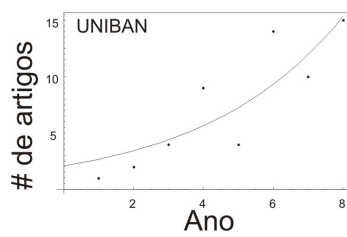
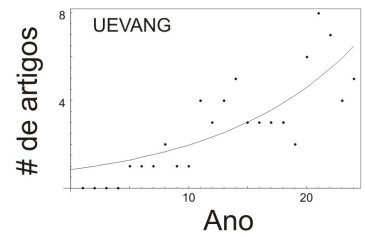
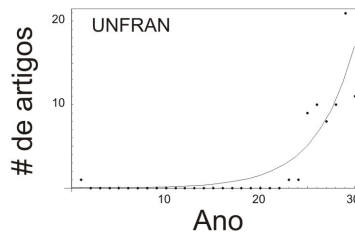
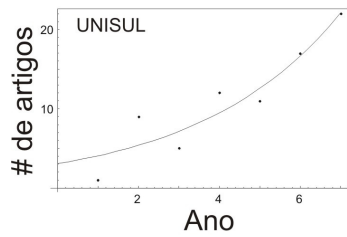
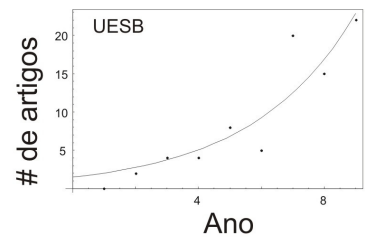
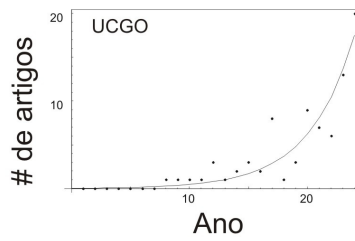
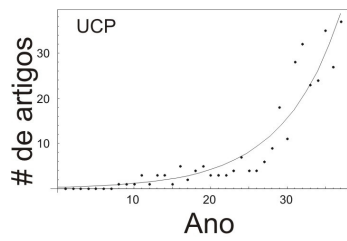












SIGLA	NOME
USP	Universidade de São Paulo
UNICAMP	Universidade Estadual de Campinas
UFRJ	Universidade Federal do Rio de Janeiro
UNESP	Universidade Estadual Paulista
UFMG	Universidade Federal de Minas Gerais
UFRGS	Universidade Federal do Rio Grande do Sul
UNIFESP	Universidade Federal de São Paulo
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
FIOCRUZ	Fundação Oswaldo Cruz
UFPE	Universidade Federal de Pernambuco
UFSCAR	Universidade Federal de São Carlos
UNB	Universidade de Brasília
UFSC	Universidade Federal de Santa Catarina
UFPR	Universidade Federal do Paraná
UFF	Universidade Federal Fluminense
UERJ	Universidade do Estado do Rio de Janeiro
PUC-RIO	Pontifícia Universidade Católica do Rio de Janeiro
CBPF	Centro Brasileiro de Pesquisas Físicas
UFBA	Universidade Federal da Bahia
UFCE	Universidade Federal do Ceará
UFV	Universidade Federal de Viçosa
UFPB	Universidade Federal da Paraíba
UEM	Universidade Estadual de Maringá
INPE	Instituto Nacional de Pesquisas Espaciais
IPEN	Instituto de Pesos e Medidas
UFSM	Universidade Federal de Santa Maria
UFRN	Universidade Federal do Rio Grande do Norte
INPA	Instituto Nacional de Pesquisas da Amazônia
UFPA	Universidade Federal do Pará
UEL	Universidade Estadual de Londrina
UFES	Universidade Estadual do Espírito Santo
PUC-RS	Pontifícia Universidade do Rio Grande do Sul
UFU	Universidade Federal de Uberlândia
UFGO	Universidade Federal de Goiás
UFRRJ	Universidade Federal Rural do Rio de Janeiro

SIGLA	NOME
UFPEL	Universidade Federal de Pelotas
OBSNAC	Observatório Nacional
IALR	Instituto Adolfo Lutz Registro
ITA	Instituto Tecnológico da Aeronáutica
UFAL	Universidade Federal de Alagoas
PETROBRAS	Petróleo Brasileiro S.A.
IMPA	Instituto Nacional de Matemática Pura e Aplicada
UFLA	Universidade Federal de Lavras
IME	Instituto Militar de Engenharia
UFJF	Universidade Federal de Juiz de Fora
IEC	Instituto Evandro Chagas
UFOP	Universidade Federal de Ouro Preto
FURG	Fundação Universidade Federal do Rio Grande
INCA	Instituto Nacional de Câncer
UFRPE	Universidade Federal Rural de Pernambuco
UFMS	Universidade Federal do Mato Grosso do Sul
FAENQUIL	Faculdade de Engenharia Química de Lorena
CEFET	Centro Federal de Educação Tecnológica
IBOT	Instituto de Botânica de São Paulo
ULBRA	Universidade Luterana do Brasil
FFFCMPA	Fundação Faculdade Federal de Ciências Médicas de Porto Alegre
IAPAR	Instituto Agrônomo do Paraná
UFSE	Universidade Federal do Sergipe
FMTM	Faculdade de Medicina do Triângulo Mineiro
UFMA	Universidade Federal do Maranhão
MPEG	Museu Paraense Emilio Goeldi
UFMT	Universidade Federal do Mato Grosso
LNCC	Laboratório Nacional de Computação Científica
UECE	Universidade Estadual do Ceará
PUC-SP	Pontifícia Universidade Católica de São Paulo
UFAM	Universidade Federal do Amazonas
UEPG	Universidade Estadual de Ponta Grossa
PUC-PR	Pontifícia Universidade Católica do Paraná
UNIVALE	Universidade Vale do Rio Doce
UMC	Universidade de Mogi das Cruzes

SIGLA	NOME
UFPI	Universidade Federal do Piauí
PUC-CAMP	Pontifícia Universidade Católica de Campinas
FMSC	Faculdade de Medicina da Santa Casa São Paulo
UDESC	Universidade do Estado de Santa Catarina
UNISINOS	Universidade do Vale do Rio dos Sinos
UNITAU	Universidade de Taubaté
UNIVAP	Universidade do Vale do Paraíba
PUC-MG	Pontifícia Universidade Católica de Minas Gerais
UPF	Universidade de Passo Fundo
UEFS	Universidade Estadual de Feira de Santana
UCB	Universidade Católica de Brasília
MACKENZIE	Universidade Presbiteria Mackenzie
UNICAXIAS	Universidade de Caxias do Sul
UESC	Universidade Estadual de Santa Cruz
UFRR	Universidade Federal de Roraima
UNIP	Universidade Paulista
UNIOESTE	Universidade Estadual do Oeste do Paraná
UCP	Universidade Católica de Petrópolis
UCGO	Universidade Católica de Goiás
UESB	Universidade Estadual do Sudoeste da Bahia
UNISUL	Universidade do Sul de Santa Catarina
UNFRAN	Universidade de Franca
UEVANG	Centro Universitário de Anápolis
UNIBAN	Universidade Bandeirante de São Paulo
UEPB	Universidade Estadual da Paraíba
UNIFOR	Universidade de Fortaleza
UNIMAR	Universidade de Marília
UNEB	Universidade do Estado da Bahia
UFAC	Universidade Federal do Acre
UNICENTRO	Universidade Estadual do Centro-Oeste
UFRO	Universidade Federal de Rondônia
USJT	Universidade São Judas Tadeu
UNOESTE	Universidade do Oeste Paulista
UEVA	Universidade Estadual Vale do Acaraú
UEMA	Universidade Estadual do Maranhão

Referências Bibliográficas

- [1] <http://portal.isiknowledge.com>
- [2] <http://www.elsevier.com>
- [3] <http://www.scopus.com>
- [4] <http://vnweb.hwwilsonweb.com/hww>
- [5] <http://www.engineeringvillage.com>
- [6] <http://pt.wikipedia.org/wiki>
- [7] A. W. Schwarz, S. Schwarz e R. J. W. Tijssen, Research and research impact of a technical university - A bibliometric study, *Scientometrics* **41**, 371-388 (1998).
- [8] G. Lewison, New bibliometric techniques for the evaluation of medical schools, *Scientometrics* **41**, 5-16 (1998).
- [9] K. Matia, L. A. N. Amaral, M. Luwel, H. F. Moed e H. E. Stanley, Scaling phenomena in the growth dynamics of scientific output, *Journal of the American Society for Information Science and Technology* **56**, 893-902 (2005).
- [10] W. Glänzel, J. Leta e B. Thijs, Science in Brazil. Part 1: A macro-level comparative study, *Scientometrics* **67**, 67-86 (2006).
- [11] J. Leta, W. Glänzel e B. Thijs, Science in Brazil. Part 2: Sectoral and institutional reserach profiles, *Scientometrics* **67**, 87-105 (2006).
- [12] S. Picoli Jr., R. S. Mendes, L. C. Malacarne e E. K. Lenzi, Scaling behavior in the dynamics of citations to scientific journals, *Europhysics Letters* **75**, 673-679 (2006).
- [13] S. Picoli Jr., Física Estatística dos Sistemas Complexos, *Tese de doutorado* (Departamento de Física, Universidade Estadual de Maringá, Maringá, 2007)
- [14] I. Gleria, R. Matsushita e S. Silva, Sistemas complexos, criticalidade e leis de potência. *Revista Brasileira de Ensino de Física* **26**, 99-108 (2004).
- [15] R. N. Mantegna e H. E. Stanley, Scaling behaviour in the dynamics of an economic index *Nature* **376**, 46-49 (1995).
- [16] S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner e Y. Dodge, Turbulent cascades in foreign exchange markets, *Nature* **381**, 767-770 (1996).

- [17] M. H. R. Stanley, L. A. N. Amaral, S. V. Buldyrev, S. Havlin, H. Leschhorn, M. Maass, M. A. Salinger e H. E. Stanley, Scaling behavior in the growth of companies, *Nature* **379**, 804-806 (1996).
- [18] K. Matia, D. Fu, S. V. Buldyrev, F. Pammolli, M. Riccaboni e H. E. Stanley, Statistical properties of business firms structure and growth, *Europhysics Letters* **67**, 498-503 (2004).
- [19] L. A. N. Amaral, S. V. Buldyrev, S. Havlin, H. Leschhorn, P. Maass, M. A. Salinger, H. E. Stanley e M. H. R. Stanley, Scaling behavior in economics: I. Empirical results for company growth, *Journal Physique I* **7**, 621-633 (1997).
- [20] S. V. Buldyrev, L. A. N. Amaral, S. Havlin, H. Leschhorn, P. Maass, M. A. Salinger, H. E. Stanley e M. H. R. Stanley, Scaling behavior in economics: II. Modeling company growth, *Journal Physique I* **7**, 635-650 (1997).
- [21] Y. Lee, L. A. N. Amaral, D. Canning, M. Meyer e H. E. Stanley, Universal features in the growth dynamics of complex organizations, *Physical Reviews Letters* **81**, 3275-3278 (1998).
- [22] D. Canning, L. A. N. Amaral, Y. Lee, M. Meyer e H. E. Stanley, Scaling the volatility of GDP growth rates, *Economics Letters* **60**, 335-341 (1998).
- [23] V. Plerou, L. A. N. Amaral, P. Gopikrishnan, M. Meyer e H. E. Stanley, Similarities between the growth dynamics of university research and of competitive economic activities, *Nature* **400**, 433-437 (1999).
- [24] T. H. Keitt e H. E. Stanley, Dynamics of North American breeding bird populations. *Nature* **393**, 257-260 (1998).
- [25] S. Picoli Jr., R. S. Mendes, L. C. Malacarne, Statistical properties of the circulation of magazine and newspaper, *Europhysics Letters* **72**, 865-871 (2005).
- [26] P. C. Ivanov, L. A. N. Amaral, A. L. Goldberger, S. Havlin, M. G. Rosenblum, Z. R. Struzik e H. E. Stanley, Multifractality in human heartbeats dynamics. *Nature (London)* **399**, 461-465 (1999).
- [27] Y. Ashkenazy, J. M. Hausdorff, P. C. Ivanov e H. E. Stanley, Nonlinear dynamical model of human gait *Physica A* **316**, 662-670 (2002).
- [28] D. Halliday, R. Resnick e J. Walker, *Fundamentos de Física, Vol. 3* (LTC, Rio de Janeiro, 1996).
- [29] A. Leike, Demonstration of the Exponential Decay Law Using Beer Froth *European Journal of Physics* **23**, 21-26 (2002).
- [30] <http://www.ncbi.nlm.nih.gov/Genbank/index.html>, acessado em 10/01/2007.
- [31] <http://www.ncbi.nlm.nih.gov>, acessado em 10/1/2007.
- [32] <http://netbsd.org/Documentation/software/pkg-growth.html>, acessado em 10/01/2007.

- [33] <http://www.worldbiblesociety.org/growth.html>, acessado em 10/01/2007.
- [34] P. F. Verhulst, *Correspondences Mathématiques et Physiques* **10**, 113-121 (1838).
- [35] <pt.wikipedia.org/wiki/LeideMoore>, acessado em 04/01/2007.
- [36] <http://www.census.gov/ipc/www/world.html>, acessado em 05/02/2007.
- [37] C. G. B. Demétrio, *Modelos Lineares Generalizados em Experimentação Agronômica* (ESALQ/USP, 2002).
- [38] <http://www.unaids.org/en/default.asp>, acessado em 10/01/2007.
- [39] M. Florio e S. Colautti, A Logistic Growth Law for Government Expenditures: An Explanatory Analysis, *Universita degli Studi di Milano (DEPA) Working Paper No. 13*. Disponível em SSRN: <http://ssrn.com/abstract=455702>.
- [40] M. T. C. Neto e D. H. Reinhardt, Relações entre parâmetros de crescimento do fruto da manga, *Revista Brasileira Fruticultura* **25**, 35-37 (2003).
- [41] E.K. Chacko, Y. T. N. Reddy e T. V. Ananthanarayanan, Studies on the relationship between leaf number and área and fruit development in mango (*Mangifera indica L.*). *Journal of Horticultural Science* **57**, 483-492, (1982).
- [42] S. S. Saini, R. N. Singh e G. S. Paliwal, Growth and Development of mango (*Mangifera indica L.*) Fruit, *Indian Journal Horticulture* **28**, 1-12 (1971).
- [43] T. R. Malthus, *An essay on the principle of population, as it affects the future improvement of society, with remarks on the speculations of Mrs. Godwin, M. Condorcet and Others Writers* (J. Johnson, London, 1798).
- [44] R. Pearl, The Biology of Population Growth, *The American Journal of Sociology* **35**, 403-410 (1929).
- [45] I. Roitt, J. Brostoff e D. Male, *Imunologia* (Manole, São Paulo, 2002).
- [46] V. Pareto, *Le Cours d'Économie Politique* (Macmillan, London, 1897).
- [47] M. J. Vilela, M. L. Martins, R. L. Mendes e A. A. Santos, Determinação de padrões de crescimento de células em cultura *Jornal Brasileiro de Patologia e Medicina Laboratorial* **39**, 67-72 (2003).
- [48] A. L. Barabási e R. Albert, Emergence of Scaling in Random Networks *Science* **286**, 509-512 (1999).
- [49] B. A. Carreras, D. E. Newman, I. Dobson e A. B. Poole, Evidence for Self-Organized Criticality in a Time Series of Electric Power System Blackouts *IEEE Transactions on circuits and systems I* **51**, 1733-1740 (2004).
- [50] A. Clauset e M. Young, Scale Invariance in Global Terrorism, *e-print, physics/0502014* (2005).
- [51] R. Eisberg e R. Resnick, *Física Quântica* (Campus, Rio de Janeiro, 1979).

- [52] B. L. Joiner, Living Histograms, *International Statistical Review* **43**, 339 (1975).
- [53] A. P. C. Almeida, E. Zandonade, M. M. Abrantes e J. A. Lamounier, Deficiência de ferro e anemia em crianças de Vitória, ES, *Pediatrics (São Paulo)* **26**, 140-150 (2004).
- [54] N. K. Razumovsky, Distribution of metal values in ore deposits, *Doklady Akademii* **28**, 814 (1940).
- [55] P. E. Sartwell, The distribution of incubation periods of infectious disease, *American Journal of Hygiene* **51**, 310-318 (1950).
- [56] P. E. Sartwell, The incubation period of poliomyelitis, *American Journal of Public Health and the National Health* **42**, 1403-1408 (1952).
- [57] P. E. Sartwell, The Incubation Period and the Dynamics of Infectious Disease, *American Journal of Epidemiology* **83**, 204-216 (1966).
- [58] K. Kondo, The log-normal distribution of the incubation time of exogenous diseases, *Japanese Journal of Human Genetic* **21**, 217-237 (1977).
- [59] A. E. Magurran, *Ecological Diversity and its Measurement* (Croom Helm, Londres, 1988).
- [60] J. Aitchison e J. A. C. Brown, *The Lognormal Distribution* (Cambridge University Press, 1957).
- [61] R.S. Mendes, L.C. Malacarne e C. Anteneodo, Statistics of football dynamics, *European Physics Journal B* **57**, 357-363 (2007).
- [62] S. R. A. Salinas, *Introdução à Física Estatística* (Editora da Universidade de São Paulo, São Paulo, 2005).
- [63] R. Gibrat, *Les Inégalités économiques*. (Sirey, Paris, 1931).
- [64] A. Király e I. M. Jánosi, Stochastic modeling of daily temperature fluctuations, *Physical Review E* **65**, 051102 (2002).
- [65] M. H. R. Stanley, S. V. Buldyrev, S. Havlin, R. Mantegna, M.A. Salinger e H. E. Stanley, Zipf plots and the size distribution of Firms, *Economic Letters* **49**, 453-457 (1995).
- [66] P. Hart e N. Oulton, Growth and Size of Firms, *The Economic Journal* **106**, 1242-1252 (1996).
- [67] S. J. Davis e J. Haltiwanger, Gross Job Creation, Gross Job Destruction, and Employment Reallocation, *Quarterly J. Economics* **107**, 819-863 (1992).
- [68] S. J. Davis, J. Haltiwanger, e S. Schuh, *Job Creation and Destruction* (MIT Press, Cambridge, Massachusetts, 1996).
- [69] A. Singh e G. Whittington, The Size and Growth of Firms, *Review Economical Studies* **42**, 15-26 (1975).

- [70] D. S. Evans, Tests of alternative theories of firm growth, *Journal of Political Economics* **95**, 657-674 (1987).
- [71] B. H. Hall, The Relationship Between Firm Size and Firm Growth in the US Manufacturing Sector, *J. Industrial Economics* **35**, 583-606 (1987).
- [72] T. Dunne, M. Roberts e L. Samuelson, The Growth and Failure of US. Manufacturing Plants, *Quarterly Journal of Economics* **104**, 671-698 (1989).
- [73] E. Garfield, Citation indexes for science - new dimension in documentation through association of ideas, *Science* **122**, 108-111 (1995).
- [74] E. Garfield, Citation analysis as a tool in journal evaluation - journals can be ranked by frequency and impact of citations for science policy studies, *Science* **178**, 471-479 (1972).
- [75] Scientific Thomson, www.scientific.thomson.com
- [76] B. B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, New York, 1977).
- [77] www.inep.gov.br/superior/censosuperior/sinopse/default.asp, acessado em 10/02/2008.
- [78] T. Tomé e M. J. Oliveira, *Dinâmica Estocástica e Irreversibilidade* (Editora da Universidade de São Paulo, São Paulo, 2001).
- [79] M. Mitzenmacher, A brief history of generative models for power law and lognormal distributions, *Internet Mathematics* **1**, 226-251 (2004).