



Universidade Estadual de Maringá
Centro de Ciências Exatas
Pós-Graduação em Física

Fernando José Antônio

SISTEMAS COMPLEXOS:
Desastre, Loteria e Epidemias

Maringá
2013



Universidade Estadual de Maringá
Centro de Ciências Exatas
Pós-Graduação em Física

Fernando José Antônio

SISTEMAS COMPLEXOS:
Desastre, Loteria e Epidemias

Tese apresentada como requisito parcial
para obtenção do título de doutor em Física,
do Programa de Pós-Graduação em Física,
da Universidade Estadual de Maringá.

Orientador:

Prof. Dr. Renio dos Santos Mendes

Maringá
2013

Dados Internacionais de Catalogação-na-Publicação (CIP)
(Biblioteca Central - UEM, Maringá – PR., Brasil)

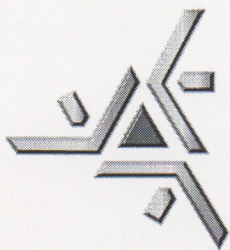
A635s Antônio, Fernando José
 Sistemas complexos: desastre, loteria e epidemias
 / Fernando José Antônio. -- Maringá, 2013.
 93 f. : il. col., figs.

 Orientador: Prof. Dr. Renio dos Santos Mendes.
 Tese (doutorado) - Universidade Estadual de
Maringá, Centro de Ciências Exatas, Programa de Pós-
Graduação em Física, 2013.

 1. Sistemas complexos. 2. Análise de dados -
Desastre - Loteria - Epidemias. Distribuição de
probabilidade. I. Mendes, Renio dos Santos, orient.
II. Universidade Estadual de Maringá. Centro de
Ciências Exatas. Programa de Pós-Graduação em Física
III. Título.

CDD 21.ed.530.13

ECSL-001268



UNIVERSIDADE ESTADUAL DE MARINGÁ

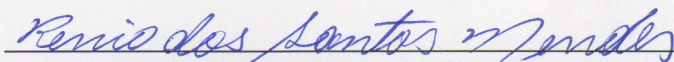
Programa de Pós-Graduação em Física

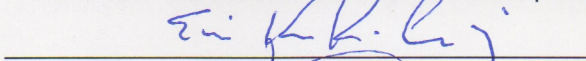
FERNANDO JOSÉ ANTÔNIO

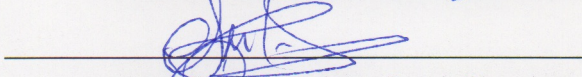
SISTEMAS COMPLEXOS: DESASTRE, LOTERIA E EPIDEMIAS

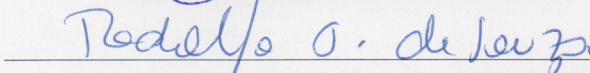
Este exemplar compreende a redação
final da tese de doutorado defendida
pelo Fernando José Antônio.

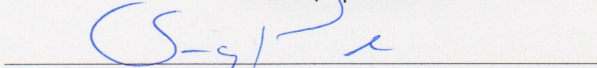
Maringá, 11 de novembro de 2013.
Banca Examinadora:


Prof. Dr. Renio dos Santos Mendes – presidente


Prof. Dr. Ervin Kaminski Lenzi - UEM


Prof. Dr. Manuel Simões Filho - UEL


Prof. Dr. Rodolfo Teixeira de Souza –
UTFPR/Apucarana


Prof. Dr. Sérgio de Picoli Júnior - UEM

Maringá
2013

AGRADECIMENTOS

Agradeço ao meu orientador, Prof. Renio dos Santos Mendes, pela atenção e amizade no decorrer do trabalho e pelos conhecimentos compartilhados.

Também gostaria de deixar meu agradecimento aos membros do Grupo de Sistemas Complexos, do Programa de Pós-Graduação em Física (UEM), pela assistência ao e pela colaboração no desenvolvimento da tese. Em particular, ao Prof. Sergio de Picoli Jr. e ao Dr. H. V. Ribeiro.

À secretária Akiko, ao programa de Pós-graduação em Física e ao Departamento de Física.

Ao grupo de pesquisa naXys do Département de Mathématique/Université de Namur, onde fui acolhido durante seis produtivos meses. Em particular, ao Prof. Renaud Lambiotte, Prof. Timoteo Carletti e P. Sampaio.

Ao apoio fundamental que recebi de familiares, professores e amigos e a todos que, de forma direta ou indireta, contribuíram para a elaboração desta pesquisa.

À Andreia, pela paciência e carinho.

À CAPES, pelo apoio financeiro.

“There is nothing like looking, if you want to find something [...]. You certainly usually find something, if you look, but it is not always quite the something you were after” (J. R. R. Tolkien, *The Hobbit*).

RESUMO

Esta pesquisa objetiva analisar dados para identificar padrões e leis universais, seguida da elaboração de modelos estocásticos em áreas multidisciplinares de sistemas complexos. Cada problema será abordado com diversos enfoques, condicionados, por sua vez, às características dos dados empíricos e de sua disponibilidade. Para tanto, o trabalho será dividido em cinco capítulos: o primeiro, de cunho biológico, está relacionado ao aumento da mortalidade de animais marinhos devido ao vazamento de petróleo no Golfo do México em 2010; no segundo capítulo, de âmbito social, discutimos o efeito coletivo das vendas em loterias no perfil do número de ganhadores, um assunto de relevante interesse em neurociência e que pode até mesmo ser conectado a sistemas biológicos; os capítulos 3, 4 e 5, por sua vez, são dedicados ao estudo do processo difusivo de epidemias no Brasil, como o HIV/AIDS e a dengue, problemas de grande interesse social.

Palavras-chave: Sistemas Complexos. Análise de Dados. Crescimento Exponencial. Loteria. Correlações. Modelo Estocástico. Epidemias. Distribuição de Probabilidades.

ABSTRACT

This research aims to analyze data in order to identify patterns and universal laws, followed by the development of stochastic models in multidisciplinary areas of complex systems. Each problem will be addressed with different approaches accordingly to the characteristics of the empirical data and their availability. For this purpose, this thesis will be divided into five chapters: the first one, from biology, is related to an increased mortality of marine animals due to the oil spill in the Gulf of Mexico in 2010; in the second chapter, in a social context, we discuss the collective effect of lottery sales in the profile of the number of winners, a topic of significant interest in neuroscience and that can even be connected to biological systems; the chapters 3, 4 and 5 are dedicated to the study of the diffusion of epidemics in Brazil, such as HIV/AIDS and dengue fever, problems of large social interest.

Keywords: Complex Systems. Data Analysis. Exponential Growth. Lottery. Correlations. Stochastic Model. Epidemics. Probability Distributions.

LISTA DE FIGURAS

1.1	Taxas de mortalidade diária acumuladas	16
2.1	Número de apostas nos últimos 200 sorteios	26
2.2	Tamanho do prêmio acumulado depois de pelo menos uma rolagem anterior dos últimos 1000 sorteios	28
2.3	Expoente de Hurst para as séries temporais dos ganhadores	30
2.4	Número de ganhadores	32
3.1	Novos casos do HIV/AIDS reportados no Brasil no ano 2009	38
3.2	Evolução da incidência anual do HIV/AIDS nas cidades brasileiras	39
3.3	Evolução temporal das taxas de incidência total	41
3.4	PDF das taxas de incidência nas cidades	43
3.5	Relação alométrica entre as taxas de incidência e a população	44
3.6	Momentos das taxas de incidência do HIV/AIDS nas cidades brasileiras no período de 1989 a 2009	46
4.1	Comparação dos dados empíricos com a solução analítica do modelo	55
4.2	Projeção das espécies segundo o modelo empírico	56
4.3	Comparação entre a solução analítica do modelo e a simulação via modelo de agentes	57
4.4	Teste chi-quadrado de Pearson	58
4.5	Segundos momentos	63
5.1	Evolução temporal dos casos de dengue no Brasil	69
5.2	Densidade da dengue pelo número de habitantes	70
5.3	PDF das taxas de incidência nas cidades	71
5.4	Relação alométrica entre as taxas de incidência e a população	72

C.1	Distribuição χ^2 com ν graus de liberdade	82
-----	--	----

SUMÁRIO

INTRODUÇÃO	12
1 O DESASTRE NO GOLFO DO MÉXICO	14
1.1 A NECESSIDADE DE QUANTIFICAR O IMPACTO	14
1.2 DADOS	15
1.3 MODELO EXPONENCIAL	16
1.4 DISCUSSÃO E PERSPECTIVAS	20
2 MEGA-SENA: PRÊMIO E PERSISTÊNCIA	22
2.1 UM APELO À NEUROCIÊNCIA	22
2.2 DADOS	23
2.2.1 Probabilidade de ganhar um prêmio	24
2.2.2 Quota de prêmios	24
2.2.3 Mega da Virada	25
2.2.4 Acerca das vendas	26
2.3 PRÊMIO ESPERADO PARA A <i>SENA</i> <i>versus</i> AUMENTO NAS VENDAS . .	27
2.4 EFEITO DE MEMÓRIA	29
2.5 MODELO ESTOCÁSTICO	31
2.6 DISCUSSÃO E PERSPECTIVAS	32
3 DINÂMICA ESPAÇO-TEMPORAL DO HIV/AIDS NO BRASIL	36
3.1 O HIV/AIDS NO MUNDO	36
3.2 PERFIL QUALITATIVO DO HIV/AIDS NO BRASIL	37
3.3 DADOS	37
3.4 EVOLUÇÃO TEMPORAL DAS TAXAS DE INCIDÊNCIA	40
3.5 DISTRIBUIÇÃO DAS TAXAS DE INCIDÊNCIA	43

3.6	TAXAS DE INCIDÊNCIA <i>versus</i> POPULAÇÃO	44
3.7	DINÂMICA COLETIVA DAS TAXAS DE INCIDÊNCIA	45
3.8	DISCUSSÃO E PERSPECTIVAS	46
4	MODELANDO A EPIDEMIA DO HIV/AIDS NO BRASIL	49
4.1	MODELOS MATEMÁTICOS DE EPIDEMIAS	49
4.2	SI EM UMA POPULAÇÃO NÃO-CONSTANTE	50
4.3	MODELO SI APLICADO AO HIV/AIDS NO BRASIL	54
4.3.1	Simulação do SI por um modelo de agentes	55
4.4	FLUTUAÇÕES ESTOCÁSTICAS DO MODELO	58
4.5	DISCUSSÃO E PERSPECTIVAS	64
5	DINÂMICA ESPAÇO-TEMPORAL DA DENGUE NO BRASIL	66
5.1	A DENGUE NO MUNDO	66
5.2	PERFIL DA DENGUE NO MAIOR PAÍS TROPICAL	67
5.3	DADOS	67
5.4	EVOLUÇÃO TEMPORAL: UM PADRÃO SAZONAL	68
5.5	DISTRIBUIÇÃO DAS TAXAS DE INCIDÊNCIA	69
5.6	TAXAS DE INCIDÊNCIA <i>versus</i> POPULAÇÃO	71
5.7	DISCUSSÃO E PERSPECTIVAS	72
	CONCLUSÃO	74
	APÊNDICES	76
A	CORRELAÇÕES	76
A.1	<i>DETRENDED FLUCTUATIONS ANALYSIS</i> – DFA	78
B	SÉRIES HIPERGEOMÉTRICAS	79
B.1	SÉRIE HIPERGEOMÉTRICA DE GAUSS	79
C	TESTE CHI-QUADRADO DE PEARSON	81
	REFERÊNCIAS	83

INTRODUÇÃO

De maneira geral, sistemas da natureza são comumente compostos por muitas partes, a partir das quais emergem comportamentos coletivos e, muitas vezes, padrões em diferentes escalas de magnitude. Sistemas com características desse tipo são usualmente denominados de *Sistemas Complexos*, compreendidos como uma forma de nos endereçarmos a uma grande classe de problemas envolvendo, por exemplo, sistemas físicos, biológicos e sociais, que, em geral, estão conectados a mais de uma área de interesse científico.

Desde as últimas décadas, o estudo de sistemas complexos tem despertado grande interesse da comunidade científica devido, dentre outros fatores, a essa grande interdisciplinaridade aliada, ainda, a um significativo potencial para aplicações práticas e/ou comerciais [1, 2]. Naturalmente, a crescente divulgação de bases de dados é também um ponto de incomensurável importância, uma vez que isso facilita o acesso de pesquisadores aos mais diversos sistemas e favorece uma abordagem mais quantitativa aos problemas.

O uso de conceitos e técnicas oriundas da física estatística tem possibilitado identificar e quantificar padrões nas mais diversas áreas do conhecimento, cujos dados não podem ser entendidos suficientemente por meio de métodos mais tradicionais. De modo geral, a análise de tais dados consiste em buscar identificar padrões na dinâmica do sistema com o objetivo de elaborar modelos teóricos e simulações numéricas.

É nesse sentido que direcionamos este trabalho, que busca analisar sistemas complexos nas áreas biológica e social para, então, elaborar modelos estocásticos a partir das informações obtidas em plataformas de dados. No capítulo 1, analisaremos dados referentes ao derramamento colossais de petróleo que ocorreu em 2010 no Golfo do México. A partir disso, é possível, por exemplo, mensurar o impacto biológico de um derramamento de petróleo no mar e como a biota da região responde a ele.

As loterias, outro exemplo de sistema complexo de grande interesse na literatura [3, 4], serão o tema do capítulo 2. Tais sistemas são fundamentalmente sociais, mas

também são problemas de grande interesse para áreas como a neurociência [5–7], visto que o valor do prêmio e o desejo por ele influenciam diretamente muitas pessoas e a dinâmica das vendas de bilhetes de apostas. Sistemas como as loterias e até mesmo certos sistemas biológicos são, em princípio, tipicamente aleatórios. Entretanto, o conjunto de todas as variáveis pode gerar um mecanismo de *feedback* capaz de introduzir complexidade quando olhamos para um comportamento coletivo, alterando perceptivelmente o cenário tipicamente aleatório *a priori*. Dessa forma, o estudo de loterias pode fornecer informações quantitativas do comportamento humano médio diante de um incentivo financeiro, o que é extremamente difícil de mensurar.

Um último exemplo notável de sistemas complexos com forte interdisciplinaridade e que será analisado aqui é o de sistemas envolvendo epidemias fundamentados em dados empíricos. Esse assunto será o foco de estudo dos capítulos 3, 4 e 5, nos quais usaremos dados da difusão do HIV/AIDS no Brasil desde o seu primeiro caso e estudará o perfil de incidência do vírus da dengue no Brasil nos últimos 12 anos. Como será visto, ambas as doenças apresentam determinadas semelhanças, como sua distribuição de casos. Dados empíricos sobre a dinâmica real de epidemias são de suma importância para a elaboração de modelos matemáticos capazes de indicar a dinâmica que rege sua difusão em determinadas regiões.

Certamente, modelos empíricos são indispensáveis para fazermos projeções da dinâmica espaço-temporal de epidemias ou, pelo menos, para identificarmos alguns de seus aspectos mais relevantes e sugerirmos melhores estratégias de controle pandêmico¹ [8,9]. Desse modo, uma clara compreensão da dinâmica de epidemias tem um impacto sensível sobre a estabilidade da qualidade de vida de um país ou região. Por fim, os Apêndices A, B e C servirão de base dos conceitos e métodos indispensáveis para a leitura desta tese.

¹ Uma doença limitada a apenas uma determinada região é considerada uma *endemia*. Quando a doença é transmitida para outras populações, infestando mais de uma região, passa a ser uma *epidemia*. Finalmente, uma epidemia que se alastra de forma desequilibrada, espalhando-se ao redor do globo, é uma *pandemia*.

O DESASTRE NO GOLFO DO MÉXICO

Neste capítulo, vamos fazer uma análise relacionada ao vazamento de petróleo no Golfo do México que ocorreu em 2010. Conhecido como *Deepwater Horizon/BP* (*British Petroleum*) [10], o vazamento iniciou-se em 22 de abril de 2010 e permaneceu por quase três meses nas águas do maior golfo do mundo (até o dia 16 de julho). Claramente, manchas de óleo nos oceanos e nos mares prejudicam o ecossistema marinho e podem, ainda, afetar alguns de água doce, em particular, no caso do acidente *Deepwater Horizon/BP*, que gerou um derramamento de grande volume e por um longo período.

1.1. A NECESSIDADE DE QUANTIFICAR O IMPACTO

Entre os vazamentos de petróleo mais bem reportados, podemos citar o *Exxon Valdez* no Alaska [11] e o *Prestige* na Espanha [12]. Entretanto, quando comparados ao *Deepwater Horizon/BP*, esses e outros vazamentos são limitados no tempo, no espaço e no volume de petróleo derramado. A grande quantidade de petróleo que vazou a uma taxa aproximadamente constante e a longa duração do acidente causaram um grande aumento na mortalidade de vertebrados presentes na biota da região.

O impacto mais visível e imediato do petróleo sobre a vida selvagem é sua adesão aos organismos, sobretudo depois que a camada de petróleo atinge a costa. Entretanto, há outros perigos relacionados aos hidrocarbonetos aromáticos policíclicos mutagênicos e/ou cancerígenos presentes no petróleo [13] e aos efeitos subletais, como as disrupções

endócrinas ou fisiológicas [14], que, por sua vez, também afetam intensamente os seres vivos. A mancha de petróleo no Golfo do México fornece, assim, uma grande oportunidade para ecologistas, toxicologistas e fisiologistas ambientais entenderem os efeitos do petróleo sobre a biota e identificarem as respostas dos organismos mais afetados pela mancha.

1.2. DADOS

Instituições governamentais americanas como a *National Oceanic and Atmospheric Administration* (NOAA) e a *U.S. Fish and Wildlife Service* (USFWS) coletaram e disponibilizaram gratuitamente *on-line* dados referentes ao desastre [15]. Com base nesses dados, foi possível investigar dois pontos:

- (i) a identificação de padrões e modelar a resposta de vertebrados tetrápodes para esse vazamento de petróleo extremo;
- (ii) a avaliação de qual(is) grupo(s) de vertebrados tetrápodes (aves, tartarugas e mamíferos) foi(ram) o(s) mais afetado(s) pelo vazamento.

Por meio desses resultados, podemos fazer uma previsão dos impactos de vazamentos de petróleo, assim como avaliar quais grupos de animais devem ser priorizados em resgates em situações como essas.

As carcaças foram coletadas ao longo de ~ 2600 km na região costeira dos EUA correspondente ao Golfo do México. E o número de animais recolhidos foi relatado ao *Unified Area Command* pelas agências USFWS e NOAA, pelos comandos da área incidente e centros de reabilitação e por outras fontes autorizadas que operam dentro da área de impacto do acidente *Deepwater Horizon/BP*.

Tendo em vista que os resultados são fundamentados apenas em resgates realizados por essas agências, assumimos um esforço de amostragem constante. E, mesmo que tal esforço não tenha sido totalmente constante, os dados são adequados para a realização, no mínimo, do segundo objetivo (avaliar qual grupo de vertebrados é mais sensível ao petróleo), afinal, ao longo de um determinado período, o esforço foi o mesmo para cada grupo de animais.

1.3. MODELO EXPONENCIAL

O volume de petróleo derramado (P) vazou a uma taxa constante ($R_P = dP/dt$; Figura 1.1a), *a priori*. Podemos notar que o conjunto de dados fornecidos pela NOAA e USFWS não contém indicações relacionadas às forças móveis do vazamento de petróleo ou do uso de dispersantes, nem mesmo as coordenadas de onde foram encontradas as carcaças. Dadas essas características, um modelo simplificado dos impactos desses acidentes sobre a fauna pode ser proposto.

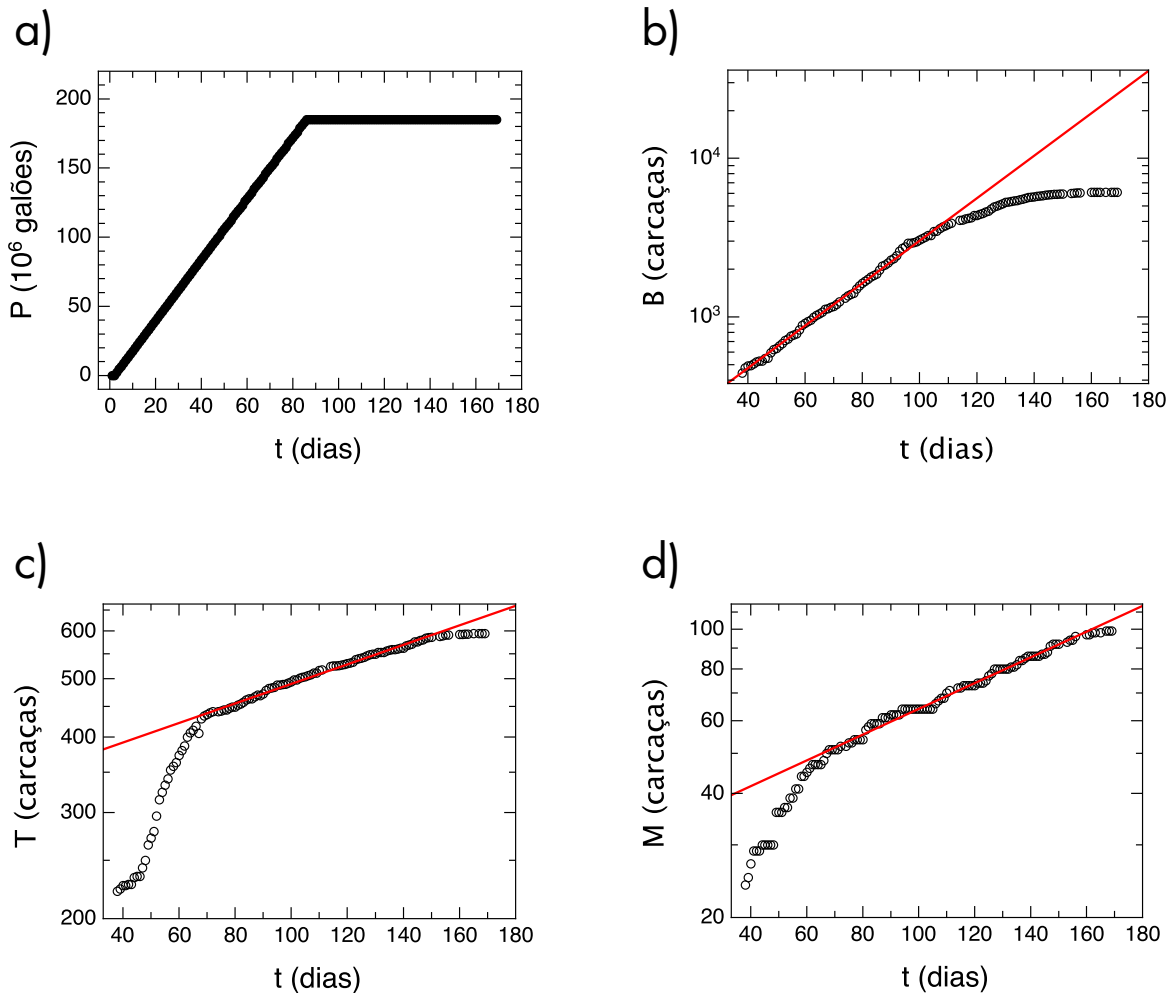


Figura 1.1: Taxas de mortalidade diária acumuladas. (a) Vazamento estimado de petróleo (P). (b) Número de carcaças de aves (B). (c) Número de carcaças de tartarugas marinhas (T). (d) Número de carcaças de mamíferos (M). As linhas sólidas são funções exponenciais ajustados aos dados.

Como consequência da mancha de petróleo, houve um crescimento na mortali-

dade de vertebrados, cujas carcaças começaram a ser reportadas 38 dias após o acidente (em 28 de Maio). Em nossa análise, consideramos os primeiros 140 dias do período de reporte de dados.

O maior impacto registrado durante esse período foi para as aves, cujo número de carcaças acumuladas (B) apresentou um crescimento exponencial (Figura 1.1b). Consequentemente, a taxa de mortalidade diária ($R_B = dB/dt$) também teve um crescimento exponencial. O crescimento intrínseco das taxas de mortalidade para esse grupo, definido como $r_B = (dR_B/dt)/R_B$, foi $r_B = 0,01338 \pm 0,00008 \, d^{-1}$. Um menor número de carcaças de tartarugas marinhas T foi reportado (Figura 1.1c). Notadamente, observamos um crescimento exponencial na mortalidade desse grupo, o que também gerou uma taxa de mortalidade diária exponencial ($R_T = dT/dt$). Entretanto, identificamos um *delay* inicial antes do estabelecimento do regime exponencial até o dia ~ 68 (27 de junho). Um número ainda menor de carcaças de mamíferos (M) foi reportado (Figura 1.1d). Para esse grupo, também houve um *delay* inicial antes do estabelecimento do regime exponencial que se manifestou a partir do dia ~ 77 (6 de julho). Os dados relacionados a mamíferos também sugerem um crescimento exponencial na mortalidade e na taxa de mortalidade diária exponencial ($R_M = dM/dt$), embora isso não seja observado tão claramente quanto no grupo das aves e no das tartarugas marinhas. O crescimento intrínseco nas taxas de mortalidade de tartarugas marinhas foi $r_T = 0,00162 \pm 0,00002 \, d^{-1}$, enquanto que o de mamíferos foi $r_M = 0,00312 \pm 0,00003 \, d^{-1}$.

A falta de um grande *delay* inicial para o grupo das aves (Figura 1.1b), tal como visto nos dois últimos grupos, pode ser justificada pelo fato de as aves serem mais sensíveis e, portanto, chegarem às praias muito antes do petróleo atingir a costa [12]. No entanto, o grupo das aves foi o primeiro a se recuperar do vazamento de petróleo, como indicado por suas taxas de mortalidade que já começaram a diminuir após o dia ~ 97 (26 de julho), enquanto que a mortalidade de tartarugas marinhas e de mamíferos ainda continuava crescendo exponencialmente até os dias ~ 155 (22 de setembro) e ~ 163 (30 de setembro), respectivamente. Essas regularidades exponenciais podem ser expressas matematicamente pela equação

$$\frac{dR_i}{dt} = r_i R_i, \quad (1.1)$$

sendo $i = B, T$ e M .

Em nossa modelagem das taxas de mortalidade de vertebrados, dois outros aspectos desejáveis foram considerados significativos:

- (i) as taxas de mortalidade R_i deveriam ser constantes quando o vazamento de petróleo parasse ($R_P = 0$);
- (ii) o aumento intrínseco na taxa de mortalidade de cada grupo de organismos deveria depender de sua sensibilidade ao petróleo.

Esses aspectos e os comportamentos exponenciais relatados são naturalmente alcançados se os aumentos intrínsecos das taxas de mortalidade forem dados por

$$r_i = a_i R_P, \quad (1.2)$$

sendo a_i relacionado apenas à sensibilidade do i -ésimo grupo de organismos ao petróleo e R_P a taxa de vazamento de petróleo. De fato, a taxa de mortalidade (R_i) para todos os grupos torna-se constante ($dR_i/dt = 0$) quando $R_P = 0$, satisfazendo o aspecto (i).

Podemos notar ainda que, quando consideramos mortes de uma população, é pertinente considerar o número de mortes por intervalo de tempo com relação ao tamanho da população. Esse aspecto pode ser considerado por meio de uma taxa de mortalidade relativa na forma

$$\check{R}_i = \frac{R_i}{P_i}, \quad (1.3)$$

em que P_i é o tamanho da população do i -ésimo grupo.

Na maior parte dos casos, tal como ocorreu com neste estudo, são disponibilizados apenas os dados sobre o número de mortes ao longo do tempo, mas não o tamanho da população. Contudo, apesar dessa falta de dados mais completos, é indispensável identificar padrões intrínsecos ao sistema. Na nossa investigação, identificamos padrões no regime em que o número de carcaças é muito menor do que o tamanho da população. Na verdade, se P_i é muito maior do que o número de mortes (mais precisamente, se $R_i \ll r_i P_i$) e, portanto, quase constante ao longo do acidente do Golfo, ele pode ser aproximado por uma constante. Nesse caso, vale a relação

$$\frac{d\check{R}_i}{dt} \simeq \frac{1}{P_i} \frac{dR_i}{dt}. \quad (1.4)$$

Consequentemente, r_i (a_i) é um parâmetro intrínseco a cada grupo, satisfazendo o

aspecto (i). Ademais, esses parâmetros já possuem uma normalização quanto ao tamanho da população, uma vez que

$$r_i = \frac{1}{R_i} \frac{dR_i}{dt} \simeq \frac{1}{\bar{R}_i} \frac{d\bar{R}_i}{dt}. \quad (1.5)$$

Portanto, o parâmetro r_i (a_i) é bastante apropriado para comparar diferentes grupos de animais, mesmo sem saber seus tamanhos populacionais. Nesse contexto, quanto maior o r_i (a_i), maior a sensibilidade do grupo i ao petróleo. Para apoiar essa suposição de independência efetiva do tamanho da população, tomemos, como exemplo, o número de carcaças de mamíferos reportados (*ca.* 100), o que representa menos de 0,15% da população de golfinhos no Golfo (*ca.* 75000) [16]. Obviamente, essa taxa é ainda menor se considerarmos o número real de mamíferos, que é muito maior, já que inclui várias outras espécies além dos golfinhos.

Um problema com os nossos dados foi o acesso limitado às taxas de mortalidade no contexto de pré-derramamento, o que nos impediu, portanto, de sabermos o quanto da mortalidade reportada está associada efetivamente à mancha de petróleo ou a outras fontes de mortalidade, sejam elas naturais ou provocadas pelo homem. Apesar de não termos dados para todos os grupos, obtivemos o número de carcaças de tartarugas encalhadas [17] e de cetáceos encalhados [18] antes do evento de poluição, para podermos comparar com os nossos dados.

De acordo com as nossas expectativas, as taxas de mortalidade de tartarugas estavam dentro da mesma ordem de magnitude antes do vazamento ($\sim 3 \text{ d}^{-1}$) e após o encerramento dele ($\sim 2 \text{ d}^{-1}$). Desse modo, é razoável supor que as elevadas taxas de mortalidade obtidas durante o derrame ($\sim 8 \text{ d}^{-1}$) entre os dias ~ 38 (28 de maio) e ~ 68 (28 de junho) estão relacionadas com os impactos do acidente. No entanto, para os cetáceos (que representam uma fração dos mamíferos), foi encontrado um cenário diferente. Por exemplo, em março de 2010 (isto é, ligeiramente antes do evento), o número de cetáceos encalhados era 3.4 vezes maior do que a média histórica registrada nesse mesmo período. Assim, considerando os cetáceos como uma boa aproximação para a presença de mamíferos, não é claro que o aumento do número de carcaças encontradas seja totalmente associado ao vazamento.

1.4. DISCUSSÃO E PERSPECTIVAS

Neste capítulo, usamos dados fornecidos pelas instituições NOAA e USFWS para avaliar o efeito de um vazamento de petróleo sobre a biota marinha. Os dados foram separados em três grupos de vertebrados tetrápodes (aves, tartarugas e mamíferos) coletados no mar e nas regiões costeiras afetados pelo vazamento de petróleo. O modelo apresentado na seção 1.3 reproduz adequadamente o crescimento exponencial na taxa de mortalidade dos grupos de animais coletados (medida como o número de carcaças recolhidos por dia). Esse comportamento exponencial é consistente com um cenário unificado para a taxa de mortalidade nos três grupos.

Em particular, concluímos que a sensibilidade ao petróleo de cada grupo é diferente, sendo as aves o grupo mais sensível, tal como indicado pelo aumento mais acentuado das taxas de mortalidade ao longo do tempo. Desse modo, elas devem ser priorizadas em serviços de resgate ($r_B \simeq 8,3 r_T$), seguidas pelos mamíferos ($r_M \simeq 1,9 r_T$) e, finalmente, pelas tartarugas. Naturalmente, tais esforços devem seguir uma dinâmica inteligente, uma vez que as aves se recuperam do impacto do desastre mais rapidamente que os outros grupos (veja a Figura 1.1b em comparação com as Figuras 1.1c e 1.1d) e, também, porque no caso das tartarugas marinhas e dos mamíferos, um aumento exponencial na mortalidade foi observado após diferentes atrasos iniciais. A figura também mostra que o grupo dos mamíferos parece ser o que sente o impacto do petróleo por mais tempo. No entanto, os dados de mortalidade pré-derramamento de mamíferos parecem indicar que o crescimento da taxa de mortalidade não é inteiramente uma consequência do vazamento.

Uma alternativa para melhorar o modelo proposto nesse capítulo é considerar o tempo de retardo inicial na mortalidade de cada grupo. Para tal, seria necessário conhecer os dados anteriores ao período de reporte de dados, mas, de antemão, sabemos que poderíamos empregar uma equação com tempo de retardo [19] relacionada com

$$\frac{dR_i(t)}{dt} = r_i R_i(t - \tau_i), \quad (1.6)$$

em que τ_i é um tempo transitório adequado e $i = B, T$ e M . Como se trata de um pro-

blema discreto, essa equação poderia ser resolvida utilizando uma aproximação de *Euler* e a propriedade de translação da transformada Z, por exemplo. Para tanto, seria imprescindível conhecer as τ_i condições iniciais de todos os grupos. Esse modelo mais detalhado deve ser capaz de justificar a região de interpolação entre a exponencial e os regimes assintóticos. Por fim, o comportamento exponencial referente à mortalidade de animais durante o vazamento de petróleo no Golfo do México e as equações que descrevem esse cenário específico poderiam ser testados em acidentes envolvendo outros poluentes.

MEGA-SENA: PRÊMIO E PERSISTÊNCIA

Neste capítulo, vamos abordar as loterias, o ramo mais famoso entre todos os jogos de azar [3, 4]. Referências às loterias datam desde a China antiga. Provavelmente durante a dinastia Han (por volta de 205 e 187 *a.C.*), foi criado um jogo de apostas chamado *Keno*, cujos recursos teriam até mesmo financiado a construção da Grande Muralha da China [20]. Já na Europa, a tradição deve ter começado ainda na Roma Antiga sob comando de *Augustus Caesar*. Mas foi somente em 1466, na cidade de Bruges (Bélgica), que uma loteria teria distribuído prêmios em dinheiro [20,21]. No Brasil, a primeira loteria foi lançada em 1784 em Vila Rica (atual Ouro Preto). E atualmente, loteria e sociedade estão tão interligadas que até mesmo aqueles que não jogam são afetados por sua existência, uma vez que o dinheiro da loteria tem financiado áreas do esporte e das artes, por exemplo [22,23].

2.1. UM APELO À NEUROCIÊNCIA

Ao longo da história, a necessidade e/ou desejo por atributos como alimento, poder, fama e riqueza tem assumido um papel central na dinâmica social [6, 24, 25]. Em muitos aspectos, adquirir qualquer um desses itens envolve uma cadeia de decisões e de escolhas intimamente relacionada a chances e/ou necessidades. Notavelmente, nossa vida pode ser vista como um processo constante de tomadas de decisões que abrangem desde escolhas corriqueiras do dia-a-dia a situações de grande importância. Uma vez que

a expectativa de ganhar um prêmio em dinheiro provoca a humanidade, a loteria e muitos outros tipos de jogos de aposta apareceram ao redor do mundo [3].

Apesar de ainda não ser claramente entendido, sabemos que os humanos adaptam o grau de esforço realizado de acordo com a magnitude de uma recompensa esperada [5, 7]. Entre as múltiplas alternativas possíveis, sempre tentamos escolher uma ação otimizada que maximize os benefícios (ou lucros, recompensas) [5, 26]. Em particular, a ambição também influencia o comportamento humano de modo similar a reconhecidos motivadores, como o alimento e o sexo [27]. Naturalmente, o dinheiro, agindo como um incentivo ou um reforçador, permite-nos atingir muitos objetivos mais facilmente, conduzindo-nos a um estado de autossuficiência [27–30].

2.2. DADOS

Os dados que exploraremos neste capítulo são mantidos pelo banco Caixa Econômica Federal e disponibilizados gratuitamente na internet [23]. O banco mantém *on-line* informações relacionadas aos sorteios, como o número de vencedores e o prêmio compartilhado entre os ganhadores. Especificamente, o período de tempo com dados varia desde o começo da Mega-Sena, em março de 1996, até o dia 15 de junho de 2013: os sorteios da Mega-Sena ocorriam uma vez por semana e, desde agosto de 2001, passaram a duas vezes por semana.

A Mega-Sena é um jogo de azar do tipo 6/60: 6 bolas numeradas no intervalo de 01 a 60 são sorteadas aleatoriamente e prêmios são concedidos aos jogadores que acertarem 6, 5 ou 4 números, conhecidos como *sena*, *quina* e *quadra*, respectivamente. Diferentemente de outros sistemas aleatórios, uma característica distintiva emergindo de loterias é que, quando não há ganhadores nas loterias em uma certa faixa de um determinado sorteio, seu prêmio relativo é rolado¹ para a mesma faixa no próximo. Historicamente, houve apenas rolagens² da *sena*, obviamente porque as chances de acertar todos os 6 números são muito menores do que as de acertar 5 ou 4 (veja a subseção 2.2.1).

¹ Do inglês *rolled-over*.

² Do inglês *rollovers*.

2.2.1. Probabilidade de ganhar um prêmio

As instruções para fazer uma aposta bem como outras informações estão presentes no verso dos bilhetes da Mega-Sena e também disponíveis on-line [31]. Os jogadores podem apostar de 6 a 15 números de um volante que cobre o intervalo de 01 a 60 e, conforme a quantidade de números acertados (4, 5, ou 6), receberão um prêmio (veja a subseção 2.2.2). O preço das apostas escala de acordo com quantos grupos de 6 números existem dentro do subconjunto selecionado. Assim, o preço varia de R\$ 2,00 para 6 números (somente 1 jogo possível) a R\$ 10.010,00 para 15 números ($15!/(6!9!) = 5.005$ jogos possíveis). Sabemos que a probabilidade de se acertar exatos i números de k escolhidos dentro de um conjunto de n opções é dada por [3]

$$P[X = i] = \binom{k}{i} \binom{n-k}{k-i} \left[\binom{n}{k} \right]^{-1}, \quad (2.1)$$

em que $\binom{m}{j} = \frac{m!}{j!(m-j)!}$.

Usando essa equação, podemos concluir que, ao fazer sorteios de 6 números, a chance de acertar a *sena*, a *quina* e a *quadra* é, respectivamente, 1 em 50.063.860, 1 em 154.518,1 e 1 em 2.332,35, enquanto que a chance de não acertar nenhum número é 1 em 1,9. Ou seja, probabilisticamente, mais de 50% dos jogadores não escolherão sequer um número que será, de fato, sorteado.

Naturalmente, embora as combinações sejam equiprováveis, certos caprichos humanos no processo de aposta podem gerar problemas mais desafiadores. Entre eles, há, por exemplo, a *seleção consciente*,³ que ocorre quando os jogadores escolhem seus “números da sorte” relacionados a preferências pessoais [32].

2.2.2. Quota de prêmios

De todo o rendimento da Mega-Sena, 54% são gastos com custeio e com financiamento de programas sociais [33]. O prêmio bruto corresponde a 46% do rendimen-

³ Do inglês *conscious selection*.

to [31, 33]. Devido ao imposto de renda subtraído desse número, o valor líquido para prêmios é, na verdade, 32,2% da arrecadação, o que será compartilhado entre os ganhadores segundo as seguintes quotas:

- ➡ 35% são distribuídos para os ganhadores da *sena*;
- ➡ 19% são distribuídos para os ganhadores da *quina*;
- ➡ 19% são distribuídos para os ganhadores da *quadra*;
- ➡ 22% são acumulados para o próximo sorteio com final 0 ou 5;
- ➡ 5% são acumulados para a *Mega da Virada* — o último sorteio de final 0 ou 5 do ano (veja a subseção 2.2.3).

2.2.3. Mega da Virada

Desde 2008, um sorteio especial com final 0 ou 5, chamado *Mega da Virada*, tem sido realizado na véspera de Ano Novo [34]. Esse sorteio é extremamente popular entre os brasileiros devido ao grande prêmio⁴ que é oferecido (veja os dois pontos posicionados muito acima dos outros na Figura 2.1). Obviamente, isso se deve principalmente às extensivas campanhas nas mídias de comunicação em massa, à eventual ocorrência de rolagens e aos 5% da quota de prêmios de cada sorteio da Mega-Sena separados para a *Mega da Virada*. No seu último sorteio, um robusto montante de R\$ 244.78 milhões foi compartilhado entre os 3 ganhadores da *sena*.

A *Mega da Virada* tem várias regalias. Seu período de apostas é tipicamente maior (apesar do mesmo preço de aposta) e, se nenhum jogador acertar todos os 6 números, o prêmio maior será compartilhado entre os jogadores que acertarem 5 números. Caso isso não aconteça, o prêmio será compartilhado entre os jogadores que acertarem 4. A menos que nenhum jogador acerte os 4 números, a *Mega da Virada* não irá acumular para o próximo sorteio. Como esperado, na história da Mega-Sena, isso nunca aconteceu. Outra diferença dos sorteios normais é quanto às cotas de prêmios para os jogadores:

- ➡ 62% são distribuídos para os ganhadores da *sena*;

⁴ Do inglês *jackpot*. Trata-se do grande prêmio de uma dada loteria, preferivelmente quando ele está acumulado ou é muito maior que os outros.

- ➡ 19% são distribuídos para os ganhadores da *quina*;
- ➡ 19% são distribuídos para os ganhadores da *quadra*.

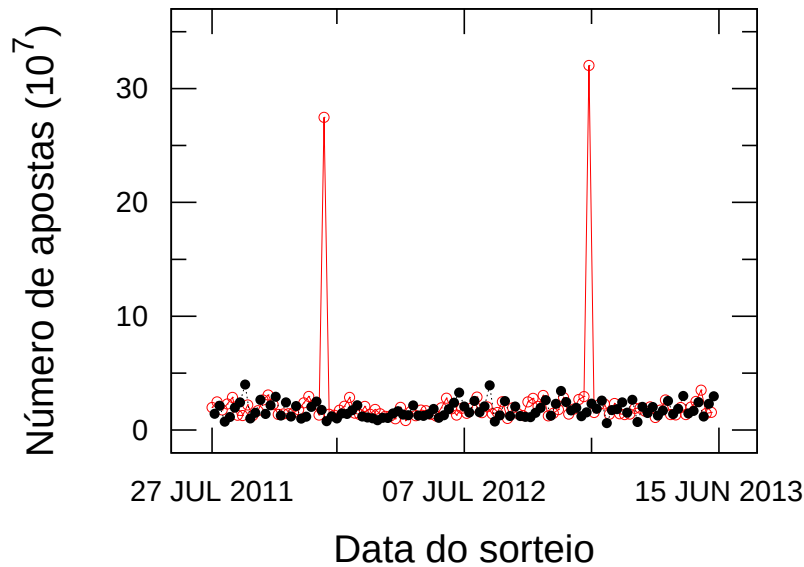


Figura 2.1: Número de apostas nos últimos 200 sorteios. Círculos cheios (●) representam o número de apostas para os sorteios das quartas-feiras, enquanto que os círculos vazios (○) representam os sorteios aos sábados.

2.2.4. Acerca das vendas

É importante notar que a discrepância entre a média de vendas para os sorteios de sábado e para os de quarta-feira, como podemos ver na Figura 2.1, é mínima, diferentemente do que acontece com a *UK Lotto*, por exemplo, cujas vendas para os sorteios de sábado são muito superiores às de quarta-feira [35]. Por esse motivo, consideramos todo os sorteios da Mega-Sena como um único conjunto de dados, portanto, independente do dia da semana. Como dito, os dois sorteios que receberam vendas significativamente maiores vistos na Figura 2.1 estão relacionados à *Mega da Virada* em 2011 e 2012. Se ignorarmos essas duas ocorrências, a diferença percentual entre as vendas médias é menor que 5%.

2.3. PRÊMIO ESPERADO PARA A SENNA VERSUS AUMENTO NAS VENDAS

Devido à oferta de grandes valores, o prêmio da *senna* é o mais cobiçado entre as loterias do Brasil. Acumulações sucessivas da *senna* induzem a um crescimento no prêmio para o próximo sorteio, conforme ilustra a Figura 2.2. Ignorando todos os sorteios com final 0 e 5 (círculos vazios na Figura 2.2), que também incorporam outros mecanismos de acumulação (veja a subseção 2.2.2 para mais detalhes), a curva principal (círculos cheios na Figura 2.2) sugere uma linha reta na forma

$$A_{t+1} = aA_t + b, \quad (2.2)$$

em que: A_t é o tamanho do prêmio para a *senna* no sorteio t , a inclinação $a = 1,12 \pm 0,01$ está relacionada ao crescimento nas vendas depois de uma rolagem e o intercepto $b = (1,71 \pm 0,38) \times 10^6$ corresponde ao valor médio dos prêmios mínimos. Esse resultado empírico indica que o tamanho do prêmio da *senna* cresce 12% depois de uma rolagem quando b puder ser negligenciado. Claramente, os dados têm flutuações menores na dinâmica de apostas, enquanto a Equação (2.2) é uma lei média determinística.

Na Figura 2.2, estamos interessados em entender como o tamanho do prêmio (relacionado ao número de bilhetes vendidos) cresce após uma rolagem. Tendo em vista que os pontos relacionados com sorteios de final 0 ou 5 estão necessariamente conectados aos seus quatro sorteios anteriores, é necessário removê-los da análise de crescimento do prêmio. Com o objetivo de obter uma comparação ainda mais realística entre preços históricos, utilizamos dados referentes à inflação no Brasil para obter a correção monetária para valores atuais [36].

Visto que o prêmio para a *senna* no sorteio $t + 1$ depende do valor acumulado do(s) sorteio(s) anterior(es) e das vendas para o mesmo sorteio, podemos determinar que

$$A_{t+1} = A_t + cN_{t+1}, \quad (2.3)$$

em que N_{t+1} é o número total de apostas para o sorteio $t + 1$ e o parâmetro $c = 0,2254$

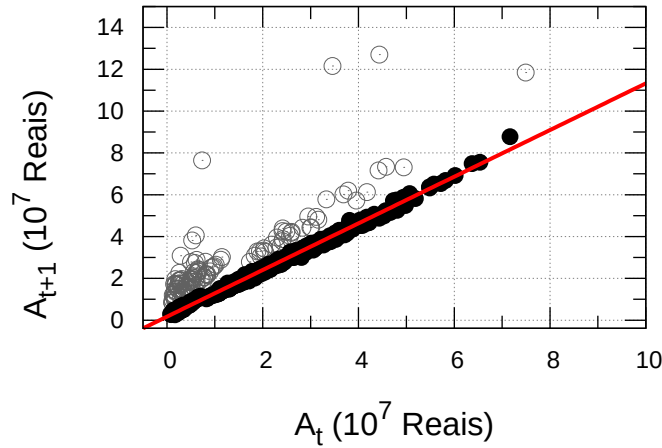


Figura 2.2: Tamanho do prêmio acumulado depois de pelo menos uma rolagem anterior dos últimos 1000 sorteios. A_{t+1} (prêmio da *sena* para o sorteio $t + 1$) como função do seu prévio valor A_t (prêmio da *sena* acumulado para o sorteio t). Os pontos relacionados aos sorteios $t + 1$ terminando em 0 ou 5 ($\circ$) desviam do comportamento coletivo principal devido a outras regras de acumulação (veja a subseção 2.2.2). Para os outros sorteios $t + 1$ ($\bullet$), a linha reta é um ajuste linear de mínimos quadrados nos dados, sugerindo a regra $A_{t+1} = 1,12A_t + 0,17$.

é o prêmio compartilhado para a *sena* multiplicado pelo custo de cada aposta (veja a subseção 2.2.2).

Evidentemente, a Figura 2.2 indica que as vendas para um sorteio específico estão intimamente relacionadas com o tamanho do prêmio previsto para *sena*. Portanto, essa figura também fornece, indiretamente, uma medida quantitativa do crescimento da demanda por bilhetes em função de um prêmio esperado.

Uma vez que a Equação (2.2) correlaciona o tamanho do prêmio no sorteio $t + 1$ com o seu valor anterior (sorteio t), o tamanho do prêmio cresce sistematicamente durante uma sucessão de acumulações. Uma sequência de acumulações para a faixa *sena* corresponde ao uso recursivo das Equações (2.2) e (2.3), conduzindo à lei exponencial

$$N_{t+1} = \frac{b}{c} a^t, \quad (2.4)$$

em que t é o número de prêmios acumulados sucessivamente para o sorteio $t + 1$. Já que $2N_{t+1}$ representa a arrecadação com as vendas para o sorteio $t + 1$, a Equação (2.4) nos informa que as vendas crescem exponencialmente em uma sucessão de acumulações e quando sorteios com final 0 e 5 não são considerados. Consistentemente com a Equação (2.4), se as vendas não fossem influenciadas pelo tamanho do prêmio acumulado,

o parâmetro a na Equação (2.2) (inclinação da Figura 2.2) seria igual a 1. Nesse caso particular, a Equação (2.4) nos revela que a média de apostas por sorteio deveria ser b/c .

2.4. EFEITO DE MEMÓRIA

Uma medida estatística caracterizando uma relação com o passado de uma série temporal é o expoente de Hurst (h). Um desvio de 0,5 desse parâmetro fornece uma medida do grau de autocorrelação de uma série temporal (veja o Apêndice A.1). Assim, o expoente de Hurst é um indicador útil de complexidade desde que um sistema passe a ser considerado mais complexo quando exibe pelo menos algum tipo de correlação. Numericamente, h pode ser calculado por meio de um método de análise fractal. Em nossa análise, usamos o DFA [37] (veja o Apêndice A.1) que, em geral, fornece uma lei de escala do tipo

$$F(s) \propto s^h, \quad (2.5)$$

na qual $F(s)$ é a flutuação quadrática média como função do tamanho s das janelas ao longo da série temporal.

Círculos cheios (•) na Figura 2.3 mostram o expoente de Hurst calculado da série temporal dos ganhadores da *sena*, da *quina* e da *quadra*, respectivamente. A ordem das tendências removidas tanto para os dados empíricos quanto para os dados simulados foi 2 para a *sena* e 3 para as demais. Naturalmente, consideramos todos os sorteios, inclusive os de final 0 e 5, para manter a continuidade dos dados. Aparentemente, nenhuma autocorrelação emerge na *sena* ($h_6 = 0,47 \pm 0,02$), mas surge um pouco na *quina* ($h_5 = 0,56 \pm 0,02$) e em uma quantidade bastante relevante na *quadra* ($h_4 = 0,68 \pm 0,02$). Esses valores aproximados foram obtidos por meio de um ajuste de mínimos quadrados à função de flutuação em escala log-log. A Figura 2.3 indica, ainda, que, “quanto mais fácil de ganhar”, mais complexa é a série temporal.

2.5. MODELO ESTOCÁSTICO

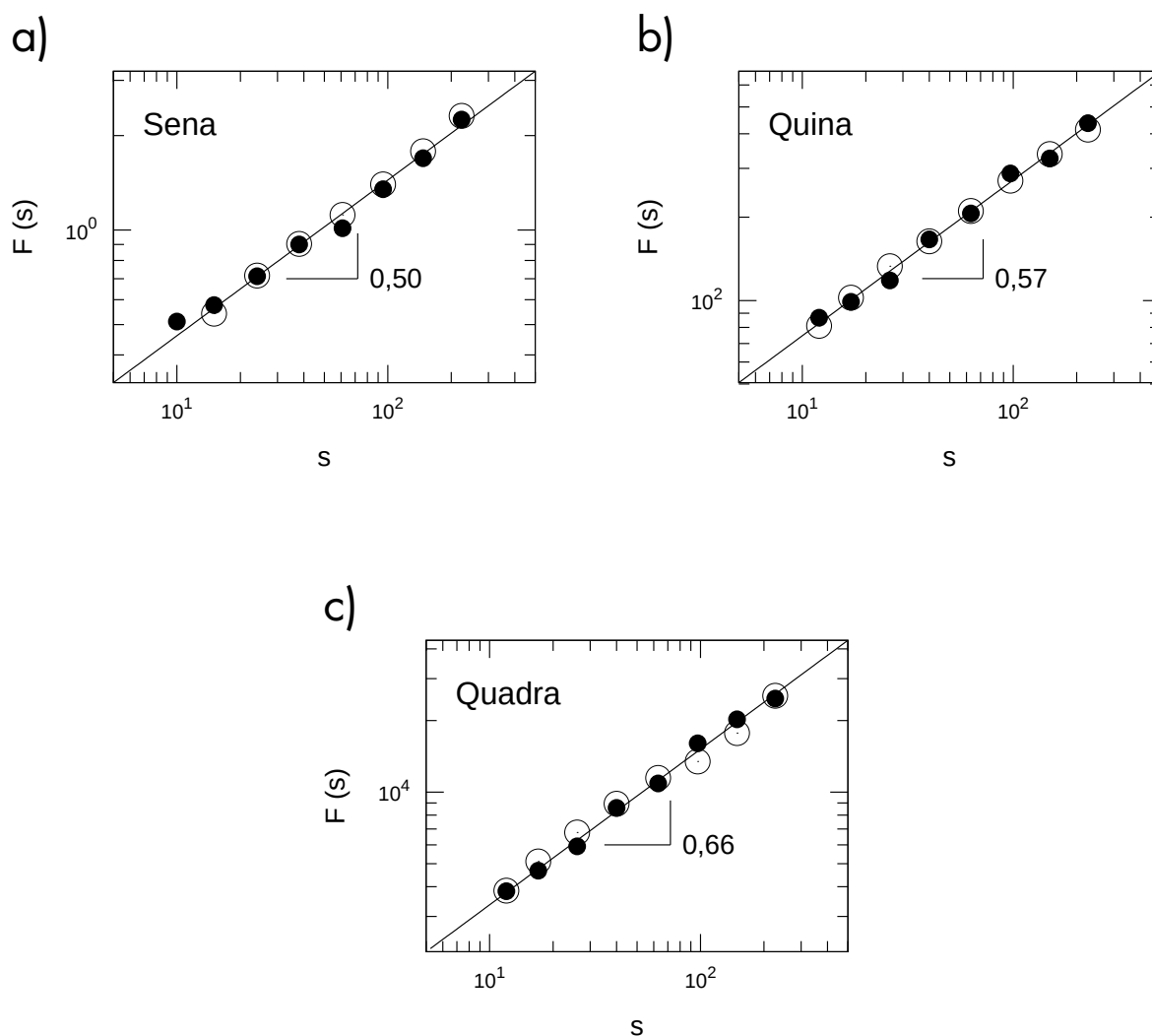


Figura 2.3: Expoente de Hurst para as séries temporais dos ganhadores. Gráfico em escala log-log da função de flutuação quadrática média $F(s)$ como função da distância linear s ao longo da série temporal para (a) *sena*, (b) *quina* e (c) *quadra*. As linhas contínuas são guias para comparar as inclinações obtidas dos dados empíricos ($\bullet$) com os dados obtidos por meio do modelo ($\circ$). A ordem das tendências removidas foi 2 para a *sena* e 3 para a *quina* e para a *quadra*.

Nesta seção, vamos apresentar um modelo para descrever a Mega-Sena. Em geral, a maior parte das loterias são variações de um único *design*:

- (i) jogadores compram bilhetes e escolhem m números diferentes de um conjunto de n opções;
- (ii) quando um sorteio ocorre, m bolas numeradas são sorteadas aleatoriamente do conjunto de n opções;
- (iii) prêmios são concedidos aos jogadores que acertam todos os m números. Em geral, jogadores que acertam alguns, mas não todos os números sorteados dividem prêmios menores.

No caso da Mega-Sena, $m = 6$ e n cobre o intervalo de 01 a 60. Visando reproduzir os padrões dos ganhadores da Mega-Sena ilustrados na Figura 2.3, propomos um modelo direto que realiza as apostas e emula os sorteios. Explicitamente, o algoritmo do modelo segue quatro regras:

- (i) realiza N apostas de 6 números diferentes selecionados aleatoriamente no intervalo de 01 a 60;
- (ii) faz um sorteio aleatório de 6 números no mesmo intervalo;
- (iii) compara as apostas com o sorteio e salva os acertos;
- (iv) se nenhuma aposta acertar os 6 números sorteados, N cresce segundo a regra empírica dada pela Equação (2.4), senão N recebe seu valor inicial N_1 .

Podemos notar que a taxa de crescimento devido à propriedade de rolagem é exatamente a que obtivemos dos dados empíricos, ou seja, $N_t = 1,12 N_{t-1}$. Já o número inicial de apostas quando não há um prêmio acumulado pode ser obtido diretamente da Equação (2.4) quando $t = 0$. Empregamos $N_1 = 3,79 \times 10^6$ apostas.

As três primeiras regras correspondem à estrutura básica de loterias em geral. Como esperado, o número de ganhadores de cada faixa, obtido de um modelo com apenas essas três regras, conduz a séries temporais que não apresentam correlações. Por outro lado, a última regra age como a pedra fundamental do nosso modelo: o número de apostas

crece exponencialmente enquanto o prêmio está rolando e ele se move de volta para seu valor inicial quando há pelo menos um ganhador do prêmio máximo. Assim, com a regra (iv) visamos introduzir o efeito motivado pelo desejo de ganhar um prêmio cada vez maior.

Uma análise fractal de autocorrelações para o resultado desse modelo está disponível na Figura 2.3, na qual os dados obtidos através do modelo ($\circ$) são comparados com os dados empíricos da Mega-Sena ($\bullet$). Com o objetivo de sobrepor as curvas, os dados do modelo foram reescalados, multiplicando um fator constante apropriado. O expoente de Hurst para o número de bilhetes premiados na *sena*, *quina* e *quadra* obtidos do modelo são $0,51 \pm 0,01$, $0,56 \pm 0,01$ e $0,64 \pm 0,02$ respectivamente, valores em bom acordo com os correspondentes obtidos dos dados empíricos. Um comparativo do número de ganhadores em cada faixa da Mega-Sena com o número obtido a partir do modelo é mostrado na Figura 2.4.

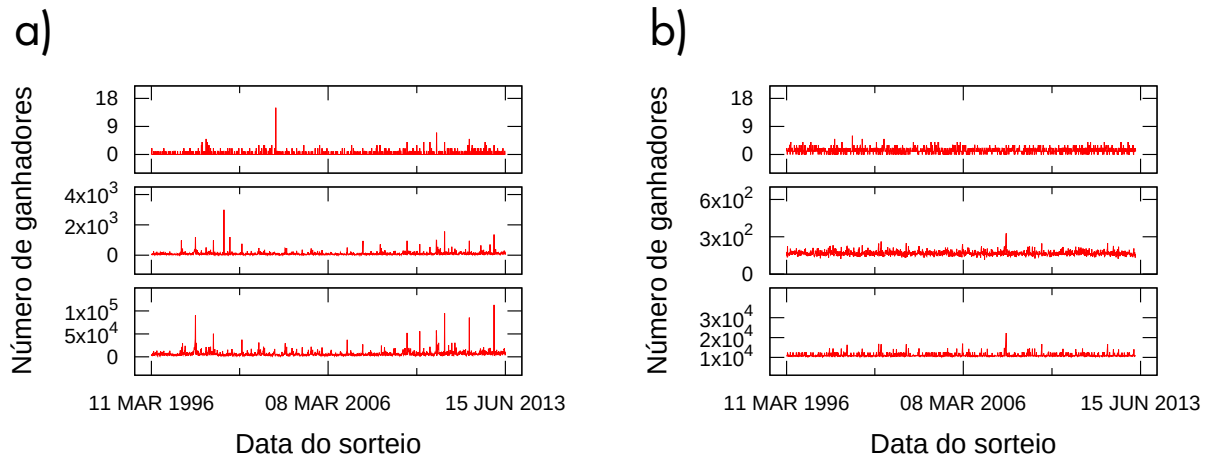


Figura 2.4: Número de ganhadores. (a) Série temporal do número de ganhadores da *sena* (acima), da *quina* (meio) e da *quadra* (abaixo). (b) Curvas análogas obtidas usando o modelo.

2.6. DISCUSSÃO E PERSPECTIVAS

Quando ocorre uma rolagem, ela se torna pública e um *jackpot* significativamente maior é esperado para a *sena* do próximo sorteio, bem como um aumento considerável na demanda por bilhetes. Quantitativamente, foi explorado na seção 2.3 uma medida ro-

busta relacionada com o aumento na compra de bilhetes por meio da relação empírica entre o tamanho do prêmio da *sena* em um sorteio $t + 1$ como função do seu valor acumulado no sorteio anterior, conforme a Equação (2.2). Esse resultado indica como a loteria, geralmente associada a um cenário aleatório composto por sorteios imparciais, pode ser influenciada por agentes externos que funcionam como um laço de *feedback* no qual as saídas do sistema o realimentam, mudando a escala da entrada. E isso pode incrementar ou decrementar a próxima saída do sistema. Claramente, a lei empírica mostrada na Figura 2.2 está, de alguma forma, relacionada com a ambição dos jogadores. Indiretamente, aquela também oferece um exemplo de como o dinheiro influencia na tomada de decisões dos humanos, um ponto clássico de interesse da neurociência [7, 38–41].

Obviamente, a Equação (2.2) foi obtida desconsiderando-se os sorteios com final 0 e 5, uma vez que eles são regidos por mais de uma rolagem (veja a subseção 2.2.2) e, portanto, apresentam um comportamento diferente, como ilustrado na Figura 2.2 (○). Por essa razão, a lei exponencial mostrada na Equação (2.4) é quebrada para os sorteios com finais 0 e 5.

Além de e por razão de o crescimento exponencial nas vendas, o número de ganhadores da Mega-Sena exibe um comportamento temporal persistente. Esse argumento pode ser justificado por meio de um modelo, que também é consistente com o fato de a Mega-Sena ser imparcial, indicando que a origem das correlações presentes nos dados é devido à ordem posicional em que as vendas aumentam. Portanto, um mecanismo externo que está conectado à Equação (2.2) deve estar subjacente a essa lembrança do passado. Claramente, podemos associar cada sorteio de um jogo de loteria com um cenário aleatório imparcial, em que muitos fatores, inclusive aqueles relacionados à ambição para o *jackpot*, geram um crescimento do número de apostas, enquanto o *jackpot* está em crescente aumento. Por esse motivo, o número de vencedores também aumenta, proporcionando um mecanismo de geração de persistência na série temporal dos ganhadores, tal como ilustrado na Figura 2.3.

À primeira vista, $h_6 \sim 0,5$ parece ser inconsistente com a suposição de que os jogadores apresentem comportamento persistente. No entanto, as chances de se ganhar um prêmio da *sena* são mínimas quando comparadas ao número médio de apostas. Como exemplo, 76,5% dos últimos 1000 sorteios não tiveram nenhum ganhador dessa faixa de prêmio. É por isso que a série temporal do número de ganhadores da *sena* tem, normal-

mente, muitos zeros e outros pequenos valores. Como consequência, o efeito do aumento no número de apostas é muito pequeno para ser perceptível, devido ao grau de aleatoriedade dos dados. Em contraste, vencedores da *quina* e da *quadra* surgem com muito mais frequência, o que torna o aumento do número de apostas mais perceptível para a medida de correlação. Além disso, as médias dos valores do expoente Hurst nas séries embaralhadas conduzem a $h_s \simeq 0,5$ tanto para os dados empíricos quanto para as séries simuladas, reforçando que as autocorrelações temporais são devidas à ordem posicional na qual as vendas aumentam. Esse teste também é consistente (em média) com sorteios imparciais e honestos ou, pelo menos, com aqueles em que as bolas não são tendenciosas, nem o nosso gerador de números aleatórios.

Como ilustrado nas Figuras 2.3 e 2.4, um modelo simples é suficiente para reproduzir o comportamento nos dados. Especificamente, o modelo se baseia na característica de rolamento para mostrar que tal mecanismo pode gerar correlações em amostras aleatórias. A diferença mais marcante entre o modelo aqui proposto e os modelos clássicos é que estes são propostos apenas para modelar a demanda por bilhetes [22, 42]. Apesar de não ser o principal objetivo do nosso modelo, ele também poderia ser utilizado para investigar medidas diretamente perceptíveis, como o aumento no número de apostas depois de acumulações. Nesse caso, todas as regras de acumulação da Mega-Sena (ver as seções 2.2.2 e 2.2.3) e certas flutuações na lei determinista da Equação (2.2) deveriam ser incorporadas ao algoritmo. Essa versão estendida pode ser de interesse para prever tendências e padrões na venda de ingressos após acumulações.

Apesar desse ser um estudo de caso, podemos esperar que padrões semelhantes à lei mostrada na Figura 2.2 e ao comportamento persistente visto na Figura 2.3 também surjam em loterias de outros países. Naturalmente, há outros fatores que atuam sobre o comportamento humano e que também poderiam conduzir a persistências similares às da Figura 2.3. Exemplos expressivos são os efeitos de publicidade, de contágio, de grupo de pares e mesmo de redes sociais [43–49]. A relevância de cada um desses aspectos sobre os resultados das loterias motiva investigações mais detalhadas.

Analogamente às loterias, esperamos que um comportamento similar possa surgir em outros sistemas sociais e biológicos [2, 50–53]. Um possível mecanismo subjacente a correlações em sistemas biológicos pode ser, por exemplo, o desejo de um animal por comida durante um processo de caça. De fato, é natural supor que um animal cace cada

vez mais motivado pela “quantidade” de sua fome (obviamente, enquanto ele tiver forças para tal). Sob esse ponto de vista, alguns processos da natureza parecem ser aleatórios, e, de fato, os são, mas em diferentes escalas. De acordo com a magnitude de alguma recompensa e/ou com o desejo para com o resultado, o grau de esforço aumenta a probabilidade de sucesso, talvez gerando algum tipo de persistência.

DINÂMICA ESPAÇO-TEMPORAL DO HIV/AIDS NO BRASIL

O espalhamento de epidemias representa uma grande ameaça para o sistema público de saúde, devido ao perigo potencial que elas representam para a sociedade [54–59]. Como consequência, epidemias e suas peculiaridades têm sido objetos de intensos estudos nas últimas décadas. Neste capítulo, vamos investigar especificamente a dinâmica espaço-temporal da epidemia do HIV/AIDS no Brasil.

3.1. O HIV/AIDS NO MUNDO

A infecção pelo Vírus da Imunodeficiência Humana (HIV) causa a Síndrome da Imunodeficiência Adquirida (AIDS) em humanos e tem sido interesse de intensos estudos desde o início da década de 1980, quando o vírus passou a espalhar-se rapidamente para todos os continentes, tornando-se uma preocupação mundial [60–66]. O lentivírus¹ HIV provoca a AIDS através de sua interação com muitos tipos de células no corpo humano e se aproveita da resposta imune do hospedeiro contra ele [67]. O vírus é transmitido principalmente através de sangue e de fluidos corporais, bem como através da transmissão vertical de mães infectadas para recém-nascidos.

O estudo epidemiológico da infecção pelo HIV por meio de dados empíricos acurados tem fornecido informações essenciais para guiar controles racionais e programas

¹ Lentivírus é um tipo de vírus com um longo período de incubação e capaz de causar uma infecção persistente que, tipicamente, resulta em uma doença progressiva, crônica e fatal.

de intervenção, assim como para monitorar a tendência espaço-temporal da epidemia em diversos países [60,63,65,66,68,69]. Há estudos mostrando que novas infecções pelo HIV entre adultos dependem de sua região geográfica e da distribuição de grupos étnicos [70].

3.2. PERFIL QUALITATIVO DO HIV/AIDS NO BRASIL

O Brasil — o quinto maior país do mundo, com quase 200 milhões de habitantes — foi o primeiro país em desenvolvimento a implementar em larga escala programas de controle e de intervenção contra o HIV/AIDS [71,72]. A resposta brasileira ao HIV/AIDS tem incluído efetivas estratégias de prevenção com forte participação da sociedade civil, mobilização multisetor e um intenso uso de terapias antirretrovirais [64,73]. Em um cenário global, tais fatos têm atraído significativamente a atenção da comunidade científica para a dinâmica do HIV/AIDS no Brasil [69,74].

Estimativas da Organização Mundial da Saúde (WHO) mostram que, no final de 2010, mais de 30 milhões de indivíduos viviam com a infecção no mundo. A título de comparação, cerca de 2,7 milhões de novas infecções foram reportadas no mundo somente nesse mesmo ano [69,72]. Esses números revelam que cerca de 1/3 da população infectada com o HIV nas Américas Central e do Sul está centrada no Brasil, o maior país em território e número de habitantes dessa região.

3.3. DADOS

Os dados sobre HIV/AIDS que analisaremos foram obtidos diretamente do DATASUS, um banco de dados público fornecido pelo Sistema Único de Saúde e mantido *online* pelo Ministério da Saúde [75]. Especificamente, analisaremos as taxas de incidência anual y , definidas formalmente como o número de novos casos de HIV/AIDS diagnosticados por ano em uma determinada cidade. Ao todo, analisaremos dados no intervalo de 1980 a 2009 para uma soma de 4901 cidades brasileiras com pelo menos um caso de infecção pelo HIV/AIDS nesse período.

Os dados disponibilizados cobrem todos os casos de infecção pelo HIV/AIDS identificados no Brasil. O primeiro caso foi identificado em 1980 na cidade de São Paulo. Nos anos seguintes, novos casos satélites foram identificados e a infecção começou a se espalhar para uma área cada vez maior.

O estado atual da infecção pelo HIV/AIDS no Brasil é ilustrado na Figura 3.1, na qual o tamanho dos círculos e a sua respectiva cor é proporcional ao logaritmo da incidência anual, y , para cada cidade brasileira. A evolução da incidência de HIV/AIDS entre as cidades brasileiras pode ser vista na Figura 3.2, que mostra (em escala logarítmica) a difusão espaço-temporal da doença no país desde o seu primeiro foco. Conforme ilustrado nas Figuras 3.1 e 3.2, a infecção está concentrada em algumas cidades-chave localizadas principalmente no sudeste e próximas ao litoral do Atlântico, onde vive a maior parte da população brasileira.

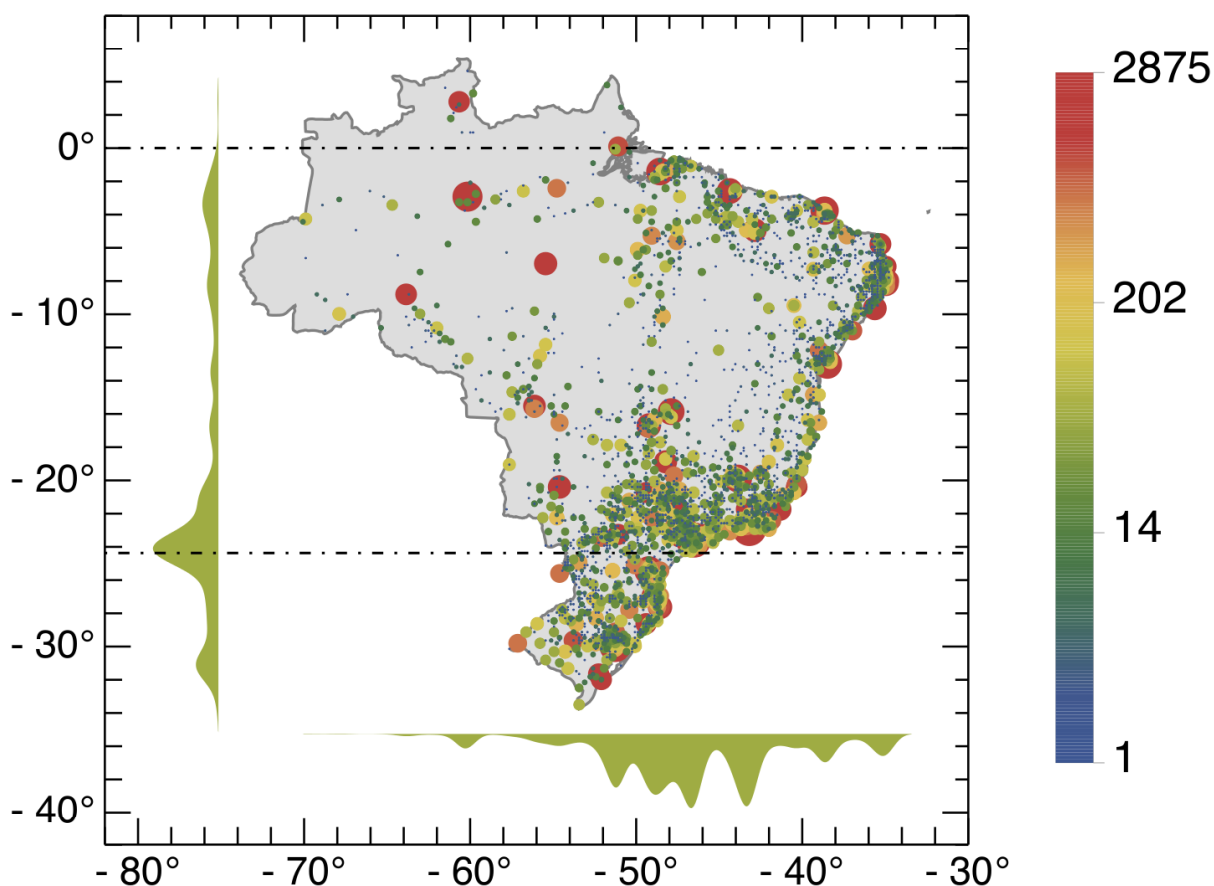


Figura 3.1: Novos casos do HIV/AIDS reportados no Brasil no ano 2009. O raio dos círculos cheios é proporcional ao logaritmo do número de casos na cidade.

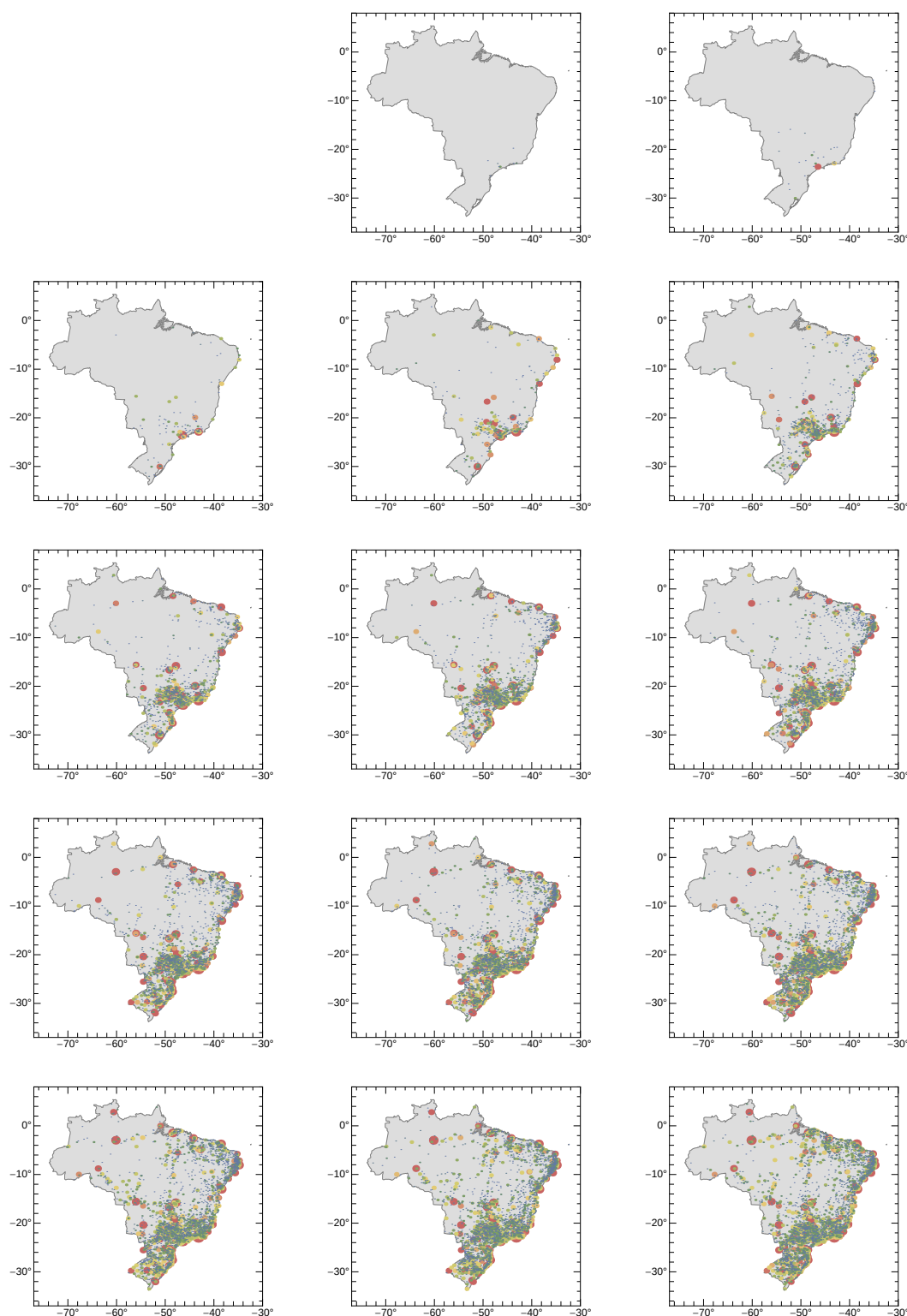


Figura 3.2: Evolução da incidência anual do HIV/AIDS nas cidades brasileiras. As figuras representam a evolução do número de casos do HIV no período de 1980 a 2008 a cada dois anos (da esquerda para a direita; de cima para baixo).²

² Usando o Adobe® Reader®, ao clicar na primeira figura será exibida uma animação da evolução da incidência anual do HIV/AIDS no Brasil de 1980 a 2009.

3.4. EVOLUÇÃO TEMPORAL DAS TAXAS DE INCIDÊNCIA

Inicialmente, consideramos Y a taxa de incidência total para o t -ésimo ano, obtida a partir da soma das taxas y das $M = 4901$ cidades brasileiras com pelo menos um caso da doença identificado. Assim,

$$Y = \sum_{i=1}^M y_i. \quad (3.1)$$

O gráfico dessa quantidade como função do tempo pode ser encontrado na Figura 3.3a. Nessa mesma figura, os dados empíricos são comparados com uma curva do tipo

$$Y(t) = \frac{y_0}{1 + (b - 1)e^{-\beta t}}, \quad (3.2)$$

indicando que Y exibe um comportamento do tipo logístico no período considerado. Um ajuste aos dados sugere os parâmetros $y_0 = 34.796 \pm 530$, $b = 138 \pm 37$ e $\beta = 0,35 \pm 0,02$. Curvas como essa já foram identificadas em simulações de surtos de epidemias usando modelos do tipo SI em redes livres de escala com estruturas locais [76, 77]. Classicamente, equações do tipo logístico são vastamente aplicadas em sistemas relacionados à dinâmica de populações nas quais há uma forte competição entre fatores de crescimento e de estabilização [19], exatamente como no caso do espalhamento de doenças.

Para o caso específico do espalhamento do HIV/AIDS, o estabelecimento de um crescimento praticamente constante no número de novas infecções depois de 1999 (Figura 3.3a) pode estar associado a fatores como o impacto do crescimento no uso de terapias ARV, bem como às extensivas campanhas de prevenção realizadas no Brasil [78].

A figura 3.3b mostra Y_c , as taxas de incidências acumuladas correspondentes, definidas como

$$Y_c = \sum_{j=1}^t Y_j, \quad (3.3)$$

em que $j = 1$ corresponde a 1980.

Para tempos pequenos, $Y_c(t)$ exibe um crescimento exponencial $Y_c(t) \propto \exp(rt)$ (veja o *inset* na Figura 3.3b). Nesse cenário, r pode ser considerado uma taxa de crescimento intrínseca descrevendo o crescimento inicial da epidemia, capaz de subjacer a todo o espalhamento da infecção nas regiões ao seu redor. Naturalmente, a taxa de cres-

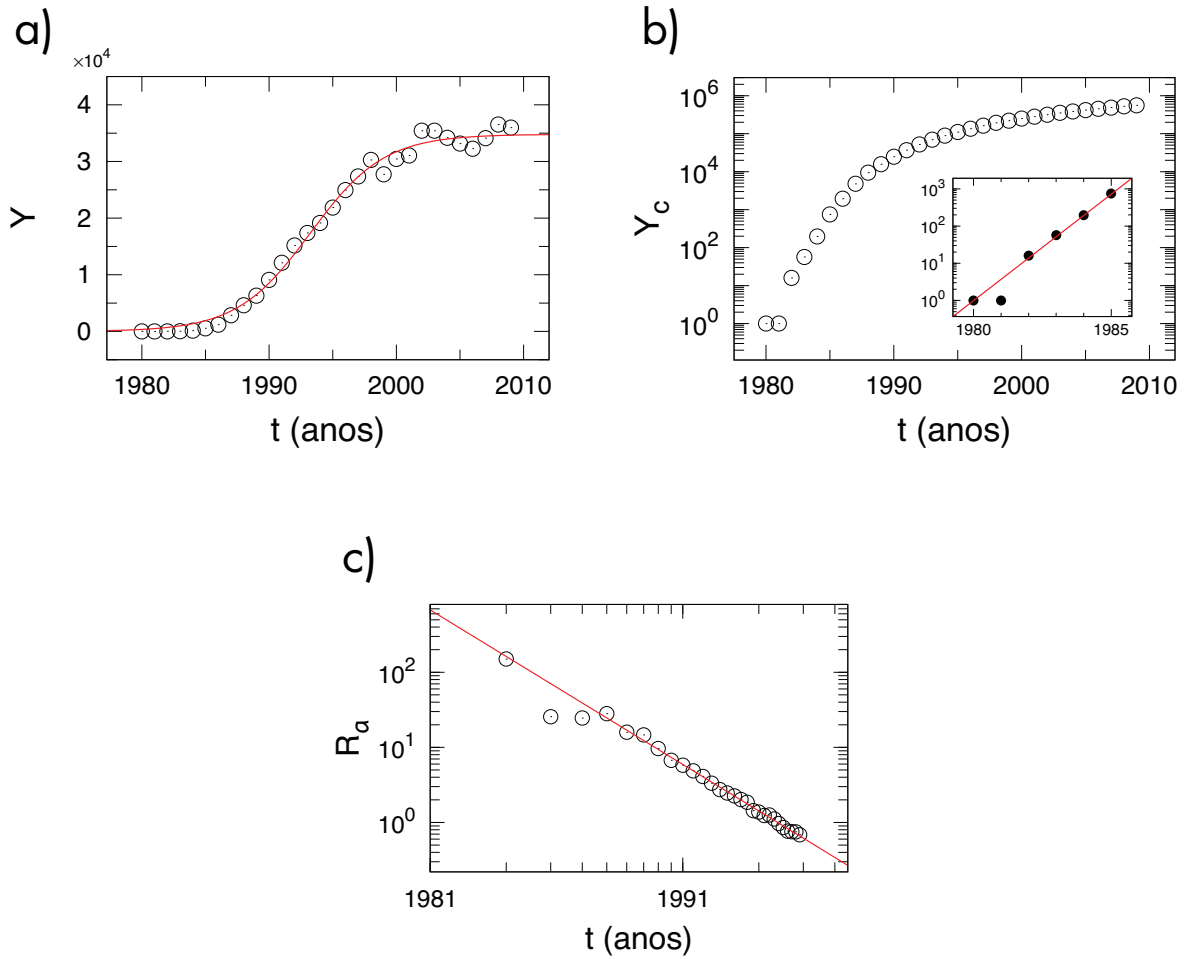


Figura 3.3: Evolução temporal das taxas de incidência total. (a) Crescimento da taxa de incidência total Y no período de 1980 a 2009 para as 4901 cidades infectadas. A linha contínua é a Equação (3.2) com os parâmetros $y_0 = 34.796$, $b = 138$ e $\beta = 0,35$. (b) Crescimento da taxa de crescimento acumulada Y_c mostrada em escala mono-log. O *inset* representa os primeiros 6 pontos da curva de Y_c , correspondendo ao período de 1980 a 1985. A linha contínua é um ajuste da equação $Y_c \propto \exp(rt)$, fornecendo o parâmetro $r = 1,32$. (c) Decaimento do número de reprodução verdadeira R_a para todo o período mostrado em escala log-log. A linha contínua é um ajuste linear, provendo o expoente $\lambda = 2,06$ para a lei de potência.

cimento intrínseca pode ser estimada por essa aproximação desde que, no início da epidemia, Y_c corresponda praticamente à prevalência da infecção pelo HIV/AIDS. Uma regressão linear dos dados conduz a $r = 1,32 \pm 0,02$ para o caso da epidemia do HIV/AIDS no Brasil. Análises similares para a epidemia do HIV/AIDS na França, no Reino Unido e na Alemanha Ocidental forneceram, respectivamente, os parâmetros $r \simeq 1,15$, $r \simeq 1,21$ e $r \simeq 2,15$ [79]. Quanto ao crescimento da epidemia do HIV/AIDS entre a população homossexual da Inglaterra e no País de Gales no período de 1981 a 1996, foi encontrado $r \simeq 0,9$ [80].

Em contraste com a taxa de crescimento intrínseca r , números de reprodução descrevem uma quantidade mais local e individual. O número de reprodução verdadeira³ R_a é definido como o número médio de casos secundários gerados diretamente por infecção com a população infectada. Uma estimativa para R_a pode ser obtida assumindo que $R_a = (Y/Y_c)D$, sendo Y a taxa de incidência, Y_c a taxa de incidência acumulada até o ano anterior e D a duração média do período infeccioso do HIV/AIDS ($\simeq 10$ anos) [73, 80]. Números de reprodução carregam importantes indicadores como o limiar para a pandemia, já que a razão incidência-por-prevalência (Y/Y_c) pode ser associada com um índice de transmissão. A figura 3.3c mostra a evolução temporal de R_a no Brasil em comparação com um decaimento do tipo lei de potência na forma

$$R_a(t) \propto t^{-\lambda}, \quad (3.4)$$

com $\lambda \simeq 2,06 \pm 0,04$. Usando as taxas de incidência do HIV/AIDS somente entre a população homossexual na Inglaterra e no País de Gales de 1981 a 1996, um padrão evidente como esse não foi identificado [80]. Claramente, os dados usados em tal estudo são apenas uma fração da população total e têm severas limitações no tempo e no espaço que não existem nos dados brasileiros.

3.5. DISTRIBUIÇÃO DAS TAXAS DE INCIDÊNCIA

³ Do inglês *actual reproduction number*.

Conforme costuma acontecer no espalhamento genérico de epidemias, a infecção pelo HIV/AIDS teve foco em uma região geográfica específica e, então, propagou-se para uma área maior (Figura 3.2). Durante o processo difusivo, indivíduos em vários centros urbanos são infectados e o raio da infecção cresce rapidamente.

Uma medida natural caracterizando tal processo pode ser obtida através de uma função capaz de identificar o número de cidades com a mesma média de hospedeiros. Visando obter mais informações nessa direção, analisamos a função densidade de probabilidade das taxas de incidência do HIV/AIDS entre cidades brasileiras $P(y)$ para cada ano. A Figura 3.4a mostra $P(y)$ para o ano 2009 em comparação com um decaimento tipo lei de potência na forma

$$P(y) \propto y^{-\alpha}. \quad (3.5)$$

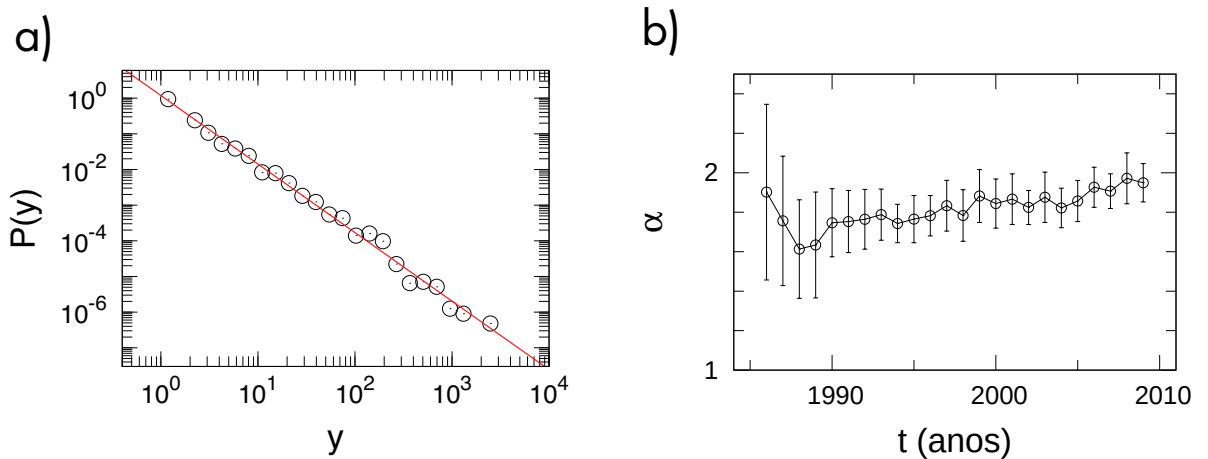


Figura 3.4: PDF das taxas de incidência nas cidades. (a) Gráfico em escala log-log da função densidade de probabilidade $P(y)$, em que y corresponde à taxa de incidência em 2009 da infecção pelo HIV em 4661 cidades brasileiras. (b) Evolução temporal do expoente α da lei de potência para o período de 1980 a 2009 e com barras de erro para 99% de confiança.

A evolução anual do expoente α é mostrada na Figura 3.4b. Em particular, encontramos o expoente $\alpha = 1,92 \pm 0,03$ para o ano 2009. Apesar das grandes flutuações nos primeiros anos (antes de 1985) devido à pequena quantidade de indivíduos infectados e considerando o intervalo de confiança de 99%, α permanece constante ao longo de todo o período ($\alpha \simeq 1,90$), sugerindo que este comportamento é robusto.

3.6. TAXAS DE INCIDÊNCIA *VERSUS* POPULAÇÃO

O número real de indivíduos infectados *per capita* depende de diferentes fatores e não deve ser constante em diferentes regiões. Assim, fatores relacionados a um comportamento de alto risco delineiam a maior parte dos novos casos. Naturalmente, a incidência de uma infecção em uma dada cidade deve estar relacionada com o seu tamanho (número de habitantes). Para investigar essa possibilidade, cruzamos os dados do HIV/AIDS com os dados referentes à população das cidades brasileiras obtidos por meio do Instituto Brasileiro de Geografia e Estatística (IBGE) — uma instituição do governo federal que realiza censos uma vez por década [81].

A Figura 3.5 mostra as taxas de incidência y do HIV/AIDS em cada cidade como uma função do seu número de habitantes p . Visando obter uma melhor aproximação, selecionamos somente as cidades com $p \geq 3 \times 10^3$ para dois anos fixados: a Figura 3.5a mostra y e p para o ano 2000, enquanto que a Figura 3.5b apresenta as taxas de incidência em cada cidade em 2009 como função de sua população p no ano 2010. As curvas obtidas são bem descritas por uma lei de escala na forma

$$y \propto p^\delta. \quad (3.6)$$

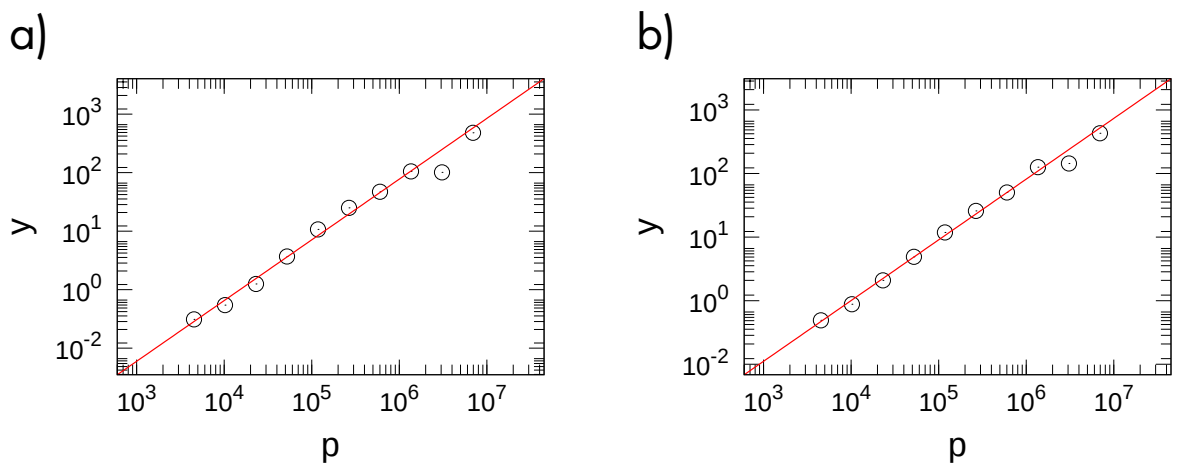


Figura 3.5: Relação alométrica entre as taxas de incidência e a população. Taxas de incidência média y (usando janelas log-espaçadas) como uma função da população p das cidades mostradas em escala log-log para anos fixados: (a) 2000 e (b) 2010. As linhas contínuas são ajustes lineares realizados para cidades com 3×10^3 ou mais habitantes. As inclinações são (a) $1,35 \pm 0,04$ e (b) $1,24 \pm 0,04$.

Particularmente, encontramos $\delta \simeq 1,35$ para o ano 2000 e $\delta \simeq 1,24$ para 2010. Resultados consistentes foram encontrados para dados do HIV/AIDS nos EUA no período de 2002 a 2003, bem como para certos indicadores sociais [82]. Notamos que as taxas de incidência da infecção pelo HIV/AIDS dependem da população com uma regra não trivial, visto que $\delta \neq 1$. Essa relação superlinear indica que a infecção cresce com a população, mostrando que, quanto mais populosa a cidade, maior sua densidade de infectados *per capita*.

Leis de escala alométricas similares às que identificamos para o HIV/AIDS também estão presentes em diversos sistemas biológicos. Exemplos clássicos são as taxas metabólicas escalando com a massa de animais e com as proporções nas plantas [83, 84].

3.7. DINÂMICA COLETIVA DAS TAXAS DE INCIDÊNCIA

Conforme já mencionado, o Brasil é uma referência mundial na elaboração de campanhas de prevenção e na concessão de serviços de tratamento [85]. O país vem investindo grande quantidade de dinheiro por vários anos. Como exemplo, somente em 2008, o governo brasileiro injetou mais de 600 milhões de dólares para esse fim, o que representa cerca de 1/3 da quantidade total investida pelo Caribe e pelas três Américas [69]. Para quantificar o impacto desses investimentos contra a incidência do HIV/AIDS, identificamos todas as cidades brasileiras com pelo menos um caso identificado do HIV desde 1989 e as separamos em quatro grupos conforme suas taxas de incidência média por ano $\langle y \rangle$:

- ➡ *Grupo I*: 4029 cidades com $\langle y \rangle < 2$;
- ➡ *Grupo II*: 565 cidades com $2 \leq \langle y \rangle \leq 10$;
- ➡ *Grupo III*: 259 cidades com $10 < \langle y \rangle \leq 84$;
- ➡ *Grupo IV*: 48 cidades com $\langle y \rangle > 84$.

Na Figura 3.6, mostramos o valor médio e a variância para cada um desses grupos. Podemos notar que as cidades pequenas (Grupos I e II), o número médio de novos casos e as respectivas flutuações ainda estão crescendo, enquanto que as grandes cidades (Grupos III e IV) exibem um comportamento diferente: o número de casos parecem ter um

viés de decrescimento, bem como suas flutuações. Isso indica que os intensos programas de combate e de controle do governo brasileiro contra a infecção têm dado resultados, sobretudo nos maiores centros urbanos, onde, também, está concentrada a maior parte dos casos.

Terapias de tratamento intensivo foram introduzidas no Brasil no fim da década de 1990 e tiveram um pico em torno de 1996 com a introdução da HAART (*Highly Active Antiretroviral Therapy*), que reduziu, substancialmente, a infectividade dos hospedeiros do HIV [64, 86]. O real impacto desses métodos sobre novos casos da infecção é evidente em cidades dos Grupos III e IV desde o final da década de 1990. Técnicas terapêuticas estão em constante investigação, visando desenvolver vacinas contra o HIV ou, pelo menos, minimizar os efeitos da infecção. Mas avanços significativos na pesquisa ainda são indispensáveis para encontrar uma solução eficiente para essa pandemia.

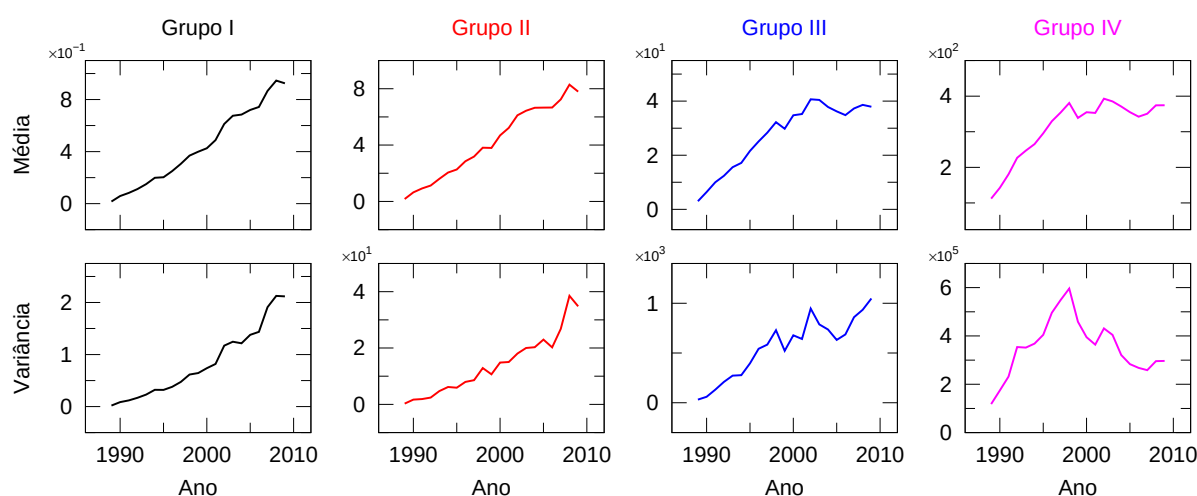


Figura 3.6: Momentos das taxas de incidência do HIV/AIDS nas cidades brasileiras no período de 1989 a 2009. O total de 4901 cidades foi separado em 4 grupos de acordo com a incidência média por ano: $\langle y \rangle < 2$ (Grupo I), $2 \leq \langle y \rangle \leq 10$ (Grupo II), $10 < \langle y \rangle \leq 84$ (Grupo III) e $\langle y \rangle > 84$ (Grupo IV).

3.8. DISCUSSÃO E PERSPECTIVAS

Neste capítulo, estudamos a evolução da epidemia do HIV/AIDS no Brasil no período de 1980 a 2009. O resultado mostrado na Figura 3.3a sugere um crescimento aproxi-

madamente logístico para a taxa de incidência total do vírus no país. Tipicamente, curvas logísticas estão associadas com uma competição entre os agentes de crescimento e de estabilização. Por essa razão, elas têm uma grande aplicação na dinâmica das populações, incluindo a disseminação de doenças infecciosas. No caso particular do HIV/AIDS, os agentes estabilizadores devem estar ligados ao impacto dos programas de controle e de intervenção contra o vírus aplicado em larga escala, como as extensas campanhas de prevenção de sua transmissão e a evolução das terapias antirretrovirais [64,85,87]. Nesse contexto, o Brasil foi o primeiro país em desenvolvimento a implementar estratégias de controle em larga escala e programas de intervenção contra a epidemia do HIV/AIDS, e a resposta brasileira a isso tem atraído grande atenção da comunidade científica.

Olhando cuidadosamente para a Figura 3.3a, podemos identificar que, após 1999, é evidente a tendência de saturação em $Y(t)$. Essa região representa o período no qual o impacto dos esforços contra o HIV/AIDS foi mais intenso, sobressaindo-se. Sabemos que uma nova infecção em uma área confinada cresce exponencialmente, logo, no caso da infecção pelo HIV/AIDS no Brasil, também encontramos esse comportamento exponencial, $\exp(rt)$, com $r = 1.3$ (*inset* da Figura 3.3b), que é semelhante ao crescimento do HIV/AIDS em alguns países europeus. Para pequenas escalas de tempo (tipicamente menores do que o tempo característico de mortes relacionadas com a infecção), a prevalência de uma infecção pode ser associada ao número total de indivíduos infectados. Usando essa suposição, obtivemos a Figura 3.3c, que mostra que a taxa global de novas infecções está rapidamente diminuindo com o tempo. Esse resultado é consistente com a literatura e prevê que a infecção está crescendo mais lentamente com o tempo. Devemos notar, ainda, que as flutuações depois de *ca.* 1992 ocorrem porque é a região na qual o intervalo de tempo pequeno já foi extrapolado (*ca.* 10 anos para o HIV/AIDS).

No intuito de explorar ainda mais a distribuição das infecções no território brasileiro, analisamos, também, como essa doença escala com a população da região onde o vírus está concentrado. Por meio da Figura 3.1, observamos que o número de casos do HIV/AIDS está bem espalhado pelo Brasil e centrado em cidades-chave, como as capitais dos estados, as regiões metropolitanas e as aglomerações sociais; já a Figura 3.2 deixa claro que a região Sudeste é o primeiro e o principal foco da doença. Em relação à população das cidades, preparamos as Figuras 3.5 e 5.4, que sugerem uma lei de escala superlinear com expoente $\delta \simeq 4/3$, uma característica típica de sistemas biológicos [83,84].

Visto que o HIV escala de modo não-trivial com a população e, segundo tudo indica, que as campanhas de combate e de controle não têm esse foco, torna-se interessante entender o impacto dos esforços contra a infecção no nível das cidades, conforme ilustrado na Figura 3.6. Usando essa figura, mostramos que os primeiros resultados dos programas e das estratégias contra a infecção pelo HIV ocorreram principalmente nos centros urbanos, com média de mais de 10 casos da doença por ano (Grupos III e IV). Em outras palavras, esse resultado sugere que a infecção não deve crescer descontroladamente em tais cidades. E, naturalmente, esse resultado é reforçado pela lei de potência decrescente para o número de reprodução verdadeira (Figura 3.3c). Por outro lado, o número médio de novas infecções por ano (bem como suas flutuações) em cidades dos grupos I e II continua crescendo. Os dados não são longos o suficiente no tempo para identificar se esse fato é apenas um período de atraso (exatamente como no caso da mortalidade de diferentes grupos de animais devido ao petróleo, descrito no capítulo 1) ou se tais cidades não foram afetadas suficientemente pelas estratégias de controle contra o HIV/AIDS.

MODELANDO A EPIDEMIA DO HIV/AIDS NO BRASIL

Neste capítulo, estudaremos novamente o crescimento na taxa de incidência do HIV/AIDS no Brasil. Nesta nova abordagem, iremos focar uma modelagem do crescimento aproximadamente logístico identificado na Figura 3.3a, por meio de conceitos de compartimentalização de indivíduos e por meio de sua modelagem através de agentes independentes.

4.1. MODELOS MATEMÁTICOS DE EPIDEMIAS

Tradicionalmente, modelos matemáticos desempenham um papel fundamental no entendimento da dinâmica de epidemias desde infecções corriqueiras pela influenza até mesmo o HIV/AIDS [88–93]. A maioria dos modelos tem base em uma compartimentalização de indivíduos de acordo com seu *status* da doença (saudável ou infectado) e seu comportamento de risco [94]. Os modelos mais simples consideram o número de indivíduos que estão susceptíveis a uma particular doença (S), o número de infectados por ela (I) e o seu número de recuperados/mortos/imunes (R). Além disso, tais modelos costumam considerar situações perfeitamente homogêneas nas quais pares de indivíduos podem interagir aleatoriamente. Assumindo, também, que os intervalos entre os eventos são distribuídos exponencialmente, esses modelos são regidos por um conjunto de equações diferenciais cuja análise proporciona uma primeira estimativa do impacto de uma doença sobre uma população.

Nos últimos anos, muitos estudos focando limitações dos modelos de campo médio foram propostos, tornando a modelagem cada vez mais real e incorporando, por exemplo, informações sobre a estrutura dinâmica da mobilidade humana. Como exemplo, temos os modelos que tiveram base em redes complexas de metapopulações,¹ que tentam simular uma estrutura social mais refinada.

Neste capítulo, vamos focar uma limitação diferente, porém, presente na maior parte dos modelos epidemiológicos: o fato de o tamanho da população não ser constante. Em particular, iremos considerar o crescimento de uma dada população ao longo do tempo que leva, por sua vez, a um crescimento no reservatório de susceptíveis e que pode ser associado a um crescimento geográfico da epidemia ou ao fluxo de indivíduos, devido à migração e/ou nascimento. Tal fenômeno pode ser desprezado no caso de epidemias de rápida evolução e fácil contágio em massa. Entretanto, ele se torna relevante quando a escala de tempo típica da epidemia puder ser comparada ao tempo de vida dos humanos, como é o caso do HIV/AIDS.

Para ilustrar o efeito do crescimento da população no espalhamento do HIV/AIDS no Brasil, consideramos um modelo SI, um modelo para epidemias que considera somente indivíduos susceptíveis ou infectados — os infectados permanecem assim até sua morte. Evidentemente, essa escolha também é motivada pelo emprego usual de modelos SI para agentes patógenos, como o HIV [96].

4.2. SI EM UMA POPULAÇÃO NÃO-CONSTANTE

Nesta seção, introduziremos o modelo estocástico por meio da formulação de uma dinâmica em nível de indivíduos para a competição de duas espécies. Uma delas são os susceptíveis (S) e a outra, os infectados (X). O primeiro grupo é uma fração da população real de uma área analisada que tem um comportamento de alto risco, o que confere a essas pessoas uma grande chance de serem infectados. Já os infectados são todos aqueles que carregam a infecção, sendo, também, denominados de *hospedeiros*.

¹ Uma metapopulação consiste em uma população formada por subpopulações ou populações locais, ou seja, trata-se de uma população espacialmente estruturada. O conceito foi introduzido por Richard Levin, em 1970, por meio de um modelo matemático elaborado para prever a dinâmica populacional de pragas de insetos em campos agrícolas [95].

Inicialmente, vamos supor que, no tempo t , a população total seja dada por $N(t) = S(t) + X(t)$ e possa ter uma dependência temporal. Conforme dito, essa é a principal diferença desse modelo para os SI tradicionais. De maneira análoga à literatura, quaisquer dois indivíduos são escolhidos aleatoriamente e interagem na mesma forma. Por essa razão, eles não possuem lembrança posicional, nem desvios do comportamento médio. Em síntese, as regras do modelo são:

- (i) associa um novo susceptível à população com probabilidade b ;
- (ii) remove um indivíduo da população com probabilidade d , independentemente de ser susceptível ou infectado;
- (iii) seleciona dois indivíduos aleatoriamente e aplica a regra de reação $S + X \rightarrow 2X$ com probabilidade λ , que é um parâmetro relacionado à taxa de transmissão da doença e à média de contatos infectivos por unidade de tempo.

Tradicionalmente, modelos relacionados a populações assumem uma densidade populacional homogênea e totalmente misturada,² ignorando quaisquer flutuações espaciais ou descrição dos indivíduos. Esse procedimento favorece a obtenção de equações determinísticas capazes de descrever a evolução temporal do sistema. Essa abordagem é chamada aproximação de campo médio,³ na qual se assume que as médias globais são boas aproximações para as distribuições locais. Devido à simplicidade das regras de reação, o modelo representa um processo *Markoviano*, que pode ser escrito matematicamente como uma *Equação Mestra* (EM) do tipo

$$\frac{\partial P(S, X, t)}{\partial t} = \sum_{(S', X') \neq (S, X)} [T(S, X | S', X') P(S', X', t) - T(S', X' | S, X) P(S, X, t)], \quad (4.1)$$

em que o primeiro termo do lado direito representa a probabilidade do sistema estar no estado (S', X') e migrar para o (S, X) no tempo t , e o segundo termo representa a probabilidade do sistema deixar o estado (S, X) no tempo t . Usando as probabilidades de

² Do inglês *fully mixed population*, trata-se de uma população na qual um indivíduo pode interagir com qualquer outro [2].

³ Do inglês *Mean-Field approach*. Devemos notar que a aproximação é chamada de *campo médio* porque as variáveis a serem observadas serão o primeiro momento das variáveis, ou seja, as médias.

transição apropriadas e predefinidas no algoritmo do modelo, a EM se torna

$$\begin{aligned}
 \frac{\partial P(S, X, t)}{\partial t} = & bP(S-1, X, t) + d \frac{(S+1)}{N} P(S+1, X, t) + \\
 & + d \frac{(X+1)}{N} P(S, X+1, t) + \\
 & + \lambda \frac{(S+1)(X-1)}{\binom{N}{2}} P(S+1, X-1, t) + \\
 & - \left[b + d \frac{S}{N} + d \frac{X}{N} + \lambda \frac{SX}{\binom{N}{2}} \right] P(S, X, t), \tag{4.2}
 \end{aligned}$$

em que o fator $\binom{N}{2}$ é um fator de correção para a escolha aleatória de pares de indivíduos por passo.

A EM fornece uma descrição completa da evolução temporal do sistema. Entretanto, ela não é simples de manipular. Um primeiro passo para atacar o problema é estudar a evolução temporal dos primeiros momentos,

$$\langle S \rangle = \sum_{S, X} S P(S, X, t) \quad \text{e} \quad \langle X \rangle = \sum_{S, X} X P(S, X, t),$$

o que conduz às seguintes equações de campo médio:

$$\begin{cases} \frac{d \langle S \rangle}{dt} = b - d \frac{\langle S \rangle}{\langle N \rangle} - 2\lambda \frac{\langle S \rangle \langle X \rangle}{\langle N \rangle^2} \\ \frac{d \langle X \rangle}{dt} = -d \frac{\langle X \rangle}{\langle N \rangle} + 2\lambda \frac{\langle S \rangle \langle X \rangle}{\langle N \rangle^2} \end{cases} \tag{4.3}$$

ao aproximarmos $\left\langle \frac{SX}{N^2 - N} \right\rangle$ por $\frac{\langle S \rangle \langle X \rangle}{\langle N \rangle^2}$, que é válida para N grande e espécies S e X não correlacionadas.

Por hipótese, $\langle N \rangle = \langle S \rangle + \langle X \rangle$ em qualquer tempo t . Então, usando as Equações (4.3), é possível escrever

$$\frac{d \langle N \rangle}{dt} = b - d, \tag{4.4}$$

cujas solução é uma função linear do tipo

$$\langle N \rangle = N_0 (1 + \gamma t), \tag{4.5}$$

com $\gamma = \frac{b-d}{N_0}$ e $N_0 = N(0)$.

A Equação (4.5) prediz um crescimento (decréscimo) linear para a população total, o que é uma aproximação adequada para a dinâmica da população brasileira desde a década de 1980. Essa equação também pode ser usada para simplificar as Equações 4.3:

$$\frac{d\phi_x}{dt} = \frac{(2\lambda - b)\phi_x - 2\lambda\phi_x^2}{N_0(1 + \gamma t)}, \quad (4.6)$$

em que $\phi_x(t)$ representa a densidade de indivíduos infectados na população N no tempo t , ou seja,

$$\phi_x = \frac{\langle X \rangle}{\langle N \rangle}. \quad (4.7)$$

Trivialmente, a densidade de agentes susceptíveis será dada por

$$\phi_s = \frac{\langle S \rangle}{\langle N \rangle} \equiv 1 - \phi_x. \quad (4.8)$$

A Equação (4.6) pode ser resolvida facilmente, conduzindo à solução

$$\phi_x(t) = \frac{(2\lambda - b)X_0}{2\lambda X_0 + [(2\lambda - b)N_0 - 2\lambda X_0](1 + \gamma t)^{-\delta}}, \quad (4.9)$$

$$\text{com } \delta = \frac{2\lambda - b}{b - d}.$$

Uma vez obtida a solução para o modelo SI proposto e buscando comparar com dados empíricos, é conveniente definir uma nova variável: $\langle Y_C \rangle$, o número médio de indivíduos infectados acumulado desde o início da epidemia na região analisada. Da Equação (4.3), $\langle Y_C \rangle$ previsto pelo modelo é dado por

$$\frac{d\langle Y_C \rangle}{dt} = 2\lambda \frac{\langle S \rangle \langle X \rangle}{\langle N \rangle^2} = 2\lambda \phi_s \phi_x. \quad (4.10)$$

Basicamente, essa equação é a soma do total de indivíduos infectados sem subtrair os mortos/removidos. Então, essa equação representa o número total de indivíduos

infectados até o tempo t desde o início da epidemia e tem como solução:

$$\begin{aligned}
 \langle Y_C \rangle = & X_0 + \left[b - \frac{2\alpha\lambda}{\delta} (\alpha\delta + \alpha - \delta) \right] t + \\
 & + \frac{2\alpha^2\lambda}{\gamma\delta} \left[\frac{1 + \gamma t}{1 + \beta(1 + \gamma t)^{-\delta}} - \frac{1}{1 + \beta} \right] + \\
 & + \frac{2\alpha\lambda}{\gamma\delta} (\alpha\delta + \alpha - \delta) (1 + \gamma t) {}_2F_1 \left[\frac{1}{1 + \frac{1}{\delta}}; -\frac{(1 + \gamma t)^\delta}{\beta} \right] + \\
 & - \frac{2\alpha\lambda}{\gamma\delta} (\alpha\delta + \alpha - \delta) {}_2F_1 \left[\frac{1}{1 + \frac{1}{\delta}}; -\frac{1}{\beta} \right], \tag{4.11}
 \end{aligned}$$

em que $\alpha = 1 - \frac{b}{2\lambda}$ e $\beta = \alpha \frac{N_0}{X_0} - 1$. Os outros parâmetros foram definidos previamente como $\gamma = \frac{b-d}{N_0}$ e $\delta = \frac{2\lambda-b}{b-d}$. Aqui, ${}_2F_1$ é uma função hipergeométrica Gaussiana (veja o Apêndice B.1).

4.3. MODELO SI APLICADO AO HIV/AIDS NO BRASIL

Nesta seção, comparamos a Equação (4.11) obtida diretamente do modelo para a curva de evolução do HIV/AIDS no Brasil. Especificamente, comparamos Y_C , o número de casos acumulados até o ano t , obtido das somas das taxas de incidência anual (y) sobre as cidades brasileiras infectadas (veja a Equação (3.1)). A taxa de incidência total Y foi definida exatamente como na seção 3.4: a soma do número de novos casos de HIV/AIDS por ano em cada cidade desde 1980 a 2009, para cada uma das 4901 cidades brasileiras com pelo menos um caso da infecção. Tomamos n como o número médio de contatos por ano ($t \rightarrow nt$) e fixamos $X_0 = 1$, já que o primeiro caso do HIV/AIDS no Brasil ocorreu em 1980 e foi o único nesse ano. Um ajuste de mínimos quadrados à Equação (4.11) para Y_C fornece os parâmetros $b = 0,201 \pm 0,053$, $d = 0,006 \pm 0,075$, $\lambda = 0,351 \pm 0,003$, $N_0 = 33.001 \pm 0$ e $n = 193.507 \pm 1.297$. Uma comparação entre dados e a solução analítica do modelo é dada na Figura 4.1.

Podemos usar, também, as soluções das equações de campo médio, Equação (4.9), para prever a força da infecção e o número de indivíduos vivos susceptíveis em qualquer tempo t . O comportamento esperado é mostrado nas Figuras 4.2a e 4.2b. Um diagrama

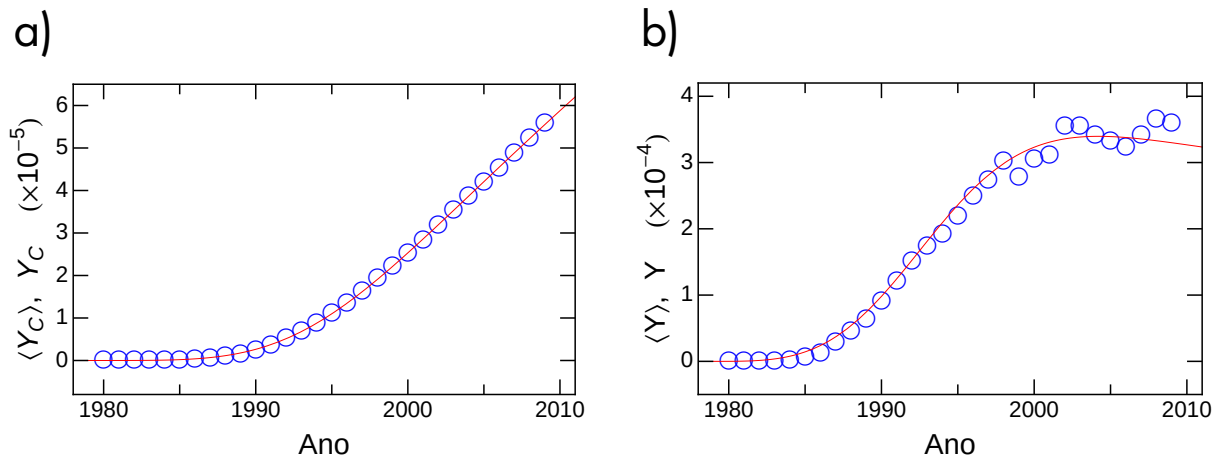


Figura 4.1: Comparação dos dados empíricos ($\circ$) com a solução analítica do modelo ($-$). (a) Força da infecção estimada pelo modelo ($\langle Y_C \rangle$; Equação (4.11)) comparada à taxa de incidência acumulada do HIV/AIDS no Brasil (Y_C ; Equação (3.3)). (b) Taxa de incidência total estimada pelo modelo ($\langle Y \rangle = \partial_t \langle Y_C \rangle$) comparada à taxa de incidência total do HIV/AIDS no Brasil (Y ; Equação (3.1)).

de fases de S e X parametrizado como função do tempo é mostrado na Figura 4.2c.

Assintoticamente, as soluções do modelo ilustradas na Figura 4.2a crescem linearmente com o tempo. Naturalmente, em tal regime, a derivada deve ir para um valor constante, como ilustrado na Figura 4.2b. Esses comportamentos mostram que o sistema atinge um estado estacionário em que o número de novas infecções (ou seja, taxa de incidência anual Y) e o número de novos susceptíveis por ano permanecem constantes.

4.3.1. Simulação do SI por um modelo de agentes

Nesta subseção, introduziremos um *Modelo de Agentes* (MA) para descrever o espalhamento do HIV por um ponto de vista microscópico. Usando o conjunto de parâmetros obtidos pelo ajuste dos dados às equações de campo médio, as regras do modelo são definidas como segue: inicialmente, definimos um conjunto contendo $N_0 - 1$ indivíduos susceptíveis e apenas um agente infectado ($X_0 = 1$). Então, em cada passo, dois agentes são selecionados aleatoriamente da população uniformemente distribuída. Se eles forem de espécies diferentes (isto é, um susceptível e um infectado), então, com uma dada probabilidade, o susceptível será infectado; senão, nada ocorre. Por outro lado, também consideramos que novos susceptíveis serão incluídos na população e que, *a priori*, as duas espécies existem na população.

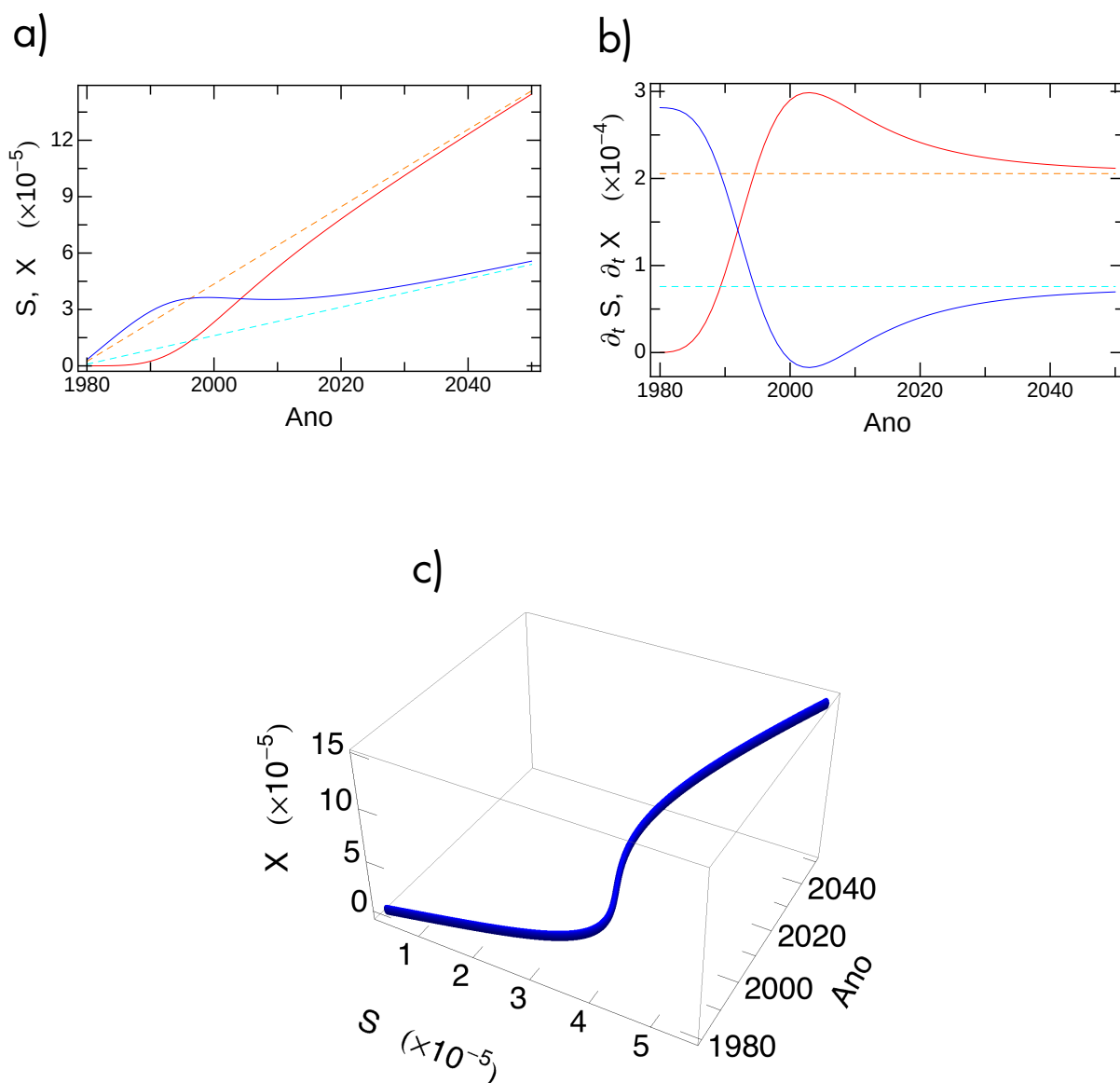


Figura 4.2: Projeção das espécies segundo o modelo empírico. (a) Número acumulado de susceptíveis (linha contínua azul) e o número acumulado de infectados (linha contínua vermelha). As linhas tracejadas correspondem ao limite assintótico. (b) Número de novos susceptíveis por ano (linha contínua azul) e o número de novos infectados por ano (linha contínua vermelha). As linhas tracejadas correspondem ao limite assintótico. (c) Diagrama de fases de S e X em função do tempo.

Para comparar o resultado do MA com sua solução analítica via campo médio (veja a Figura 4.3), simulamos numericamente o MA 1000 vezes e tomamos o comportamento médio em função do tempo.

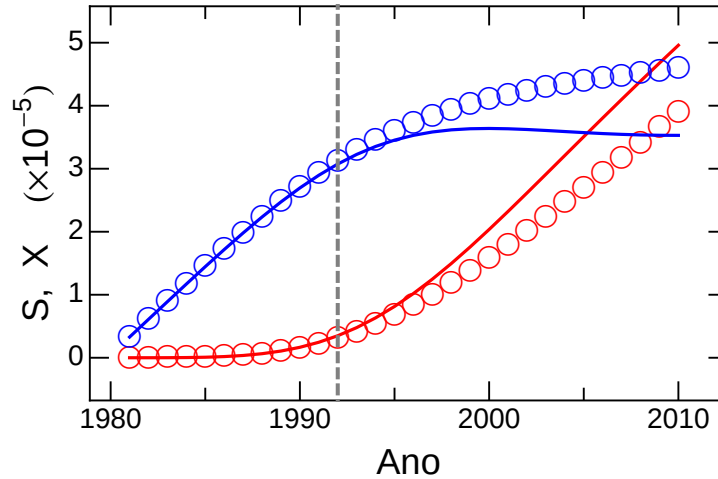


Figura 4.3: Comparação entre a solução analítica do modelo e a simulação via modelo de agentes. Número de susceptíveis (azul) e de infectados (vermelho) obtido por meio do modelo de agentes (círculos abertos) e comparado à solução analítica (linhas contínuas).

O resultado da simulação e a solução analítica estão em excelente acordo durante a fase inicial *ca.* $t \lesssim 12$ anos depois do início da infecção (linha tracejada na Figura 4.3). Entretanto, podemos notar que, à medida que o tempo cresce, um desvio aparece, mas o mesmo comportamento persiste. Quando comparada com a solução de campo médio, o número médio de agentes infectados preditos pelo MA subestima a força da infecção e sobrestima o número de susceptíveis no sistema (veja a Figura 4.3). Exemplos disponíveis na literatura sugerem que esse efeito parece estar relacionado à aproximação discreta usada no MA, a qual não existe na solução de campo médio [97, 98].

Com o intuito de entender tal desvio ao longo do tempo, calculamos os valores- p usando o teste chi-quadrado de Pearson (veja o Apêndice C). Um histograma de tais valores- p é mostrado na Figura 4.4. É notável que há um pico pronunciado em torno de 1, que ocorre quando o resultado do MA e as soluções de campo médio estão próximas.

O histograma mostra que cerca de 60% das curvas obtidas com a simulação estão na vizinhança do valor previsto teoricamente pela aproximação de campo médio (nível de significância de 0.01). Uma melhoria inicial que tem gerado muitos estudos é remover o *estado absorvente* do sistema [99, 100]. Esse procedimento consiste em manter a

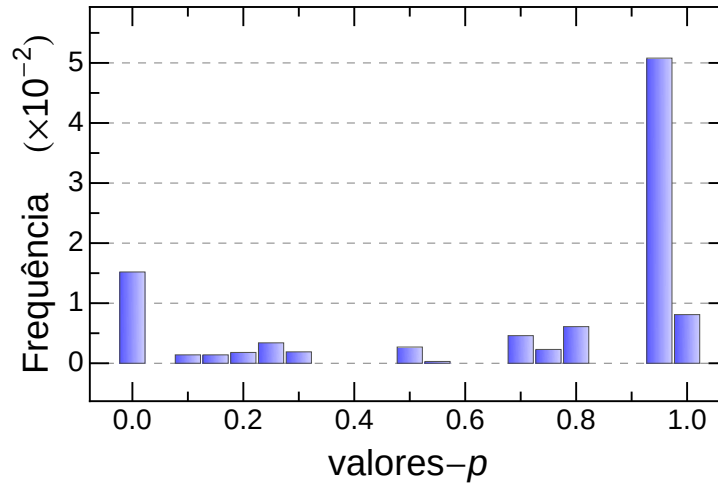


Figura 4.4: Teste chi-quadrado de Pearson. Histograma dos valores- p obtidos por meio da comparação da solução analítica da Equação (4.11) com a taxa de incidência total Y prevista pelo modelo SI.

infecção dentro da população durante todo o tempo, dessa forma, haverá sempre, pelo menos, um indivíduo infectado na população. Ainda, o primeiro pico da Figura 4.4 deverá ser extinto, visto que ele representa o estado livre da doença.

4.4. FLUTUAÇÕES ESTOCÁSTICAS DO MODELO

Nesta seção, vamos aplicar um método formal desenvolvido por van Kampen, que, além de fornecer um quadro formal para obter os mesmos resultados de uma forma mais sistemática, também fornece um método para identificar correções estocásticas em torno do estado de equilíbrio do resultado determinístico no limite de um sistema grande [101, 102]. Em outras palavras, é possível identificar o perfil das flutuações do sistema para tempos tipicamente longos. Esse procedimento vem sendo aplicado em diversas situações, como em ciclos predador-presa [103] e em reações autocatalíticas [102, 104, 105].

Para tal, vamos escrever os valores esperados para as duas espécies como $S = \sqrt{N}\xi_s + N\phi_s(t)$ e $X = \sqrt{N}\xi_x + N\phi_x(t)$, em que $N = S + X$, $\phi_s(t)$ e $\phi_x(t)$ são as soluções assintóticas e as novas variáveis ξ_s e ξ_x são os erros estocásticos supostamente aleatórios (os fatores $\sqrt{N}$ já assumem isso). Assim, podemos mostrar que $P(S, X, t) = \frac{1}{N} \Pi(\xi_s, \xi_x, t)$

e que a EM pode ser escrita como

$$\begin{aligned}
 N \frac{\partial P}{\partial t} = & (\varepsilon_s^{+1} \varepsilon_x^{-1} - 1) [T(S-1, X+1|S, X) \Pi] + \\
 & + (\varepsilon_s^{-1} - 1) [T(S+1, X|S, X) \Pi] + \\
 & + (\varepsilon_s^{+1} - 1) [T(S-1, X|S, X) \Pi] + \\
 & + (\varepsilon_x^{+1} - 1) [T(S, X-1|S, X) \Pi],
 \end{aligned} \tag{4.12}$$

em que $\varepsilon_i^{\pm k} f(n_i) = f(n_i \pm k)$ (com i indicando as variáveis S ou X) é o operador de grau⁴ proposto por van Kampen [101]. A dedução desse operador é simples: consiste em expandir $f(n_i \pm k)$ em uma série de Taylor em torno de n , o que fornece

$$f(n_i \pm k) = \sum_{j=0}^{\infty} \frac{(\pm k)^j}{j!} \left(\frac{\partial}{\partial n_i} \right)^j f(n_i). \tag{4.13}$$

Ao fazer a troca de variáveis $n_i = \sqrt{N} \xi_i + N \phi_i(t)$, obtemos

$$\varepsilon_i^{\pm k} = \sum_{j=0}^{\infty} \frac{(\pm k)^j}{j!} \left(\frac{\partial}{\partial n_i} \right)^j \implies \varepsilon_i^{\pm k} = \sum_{j=0}^{\infty} \frac{1}{j!} \left(\frac{\pm k}{\sqrt{N}} \right)^j \left(\frac{\partial}{\partial \xi_i} \right)^j. \tag{4.14}$$

Particularmente, quando $k = 1$ (veja a Equação (4.12)), temos que

$$\varepsilon_i^{\pm 1} = 1 \pm \frac{1}{\sqrt{N}} \frac{\partial}{\partial \xi_i} + \frac{1}{2N} \frac{\partial^2}{\partial \xi_i^2} + \dots \tag{4.15}$$

Note que, quando $N \gg k$ (ou seja, no limite de um sistema grande), apenas os primeiros termos da série são suficientes para escrever uma expansão. Vale ressaltar, ainda, que as derivadas parciais indicam que, se $f(\cdot)$ tiver mais de uma variável, o operador de grau agirá apenas na de mesmo nome que ele.

De um modo geral, a expansão escrita na Equação (4.12) é aplicada em grandes sistemas com tamanho constante, mas, em nosso modelo, o tamanho da população apresenta uma dependência temporal. Uma vez que ainda se trata de um sistema de grande porte, por definição, a expansão de van Kampen também poderá ser aplicada com o objetivo de identificar o perfil das flutuações em torno dos limites assintóticos. Porém, como N agora depende do tempo, novos termos devem ser considerados no lado esquerdo da

⁴ Do inglês *step operator*.

EM ao fazermos a troca das variáveis $\{S, X\}$ para $\{\xi_s, \xi_x\}$. Usando as regras estabelecidas pelo algoritmo do modelo e, em particular, a Equação (4.4), o lado esquerdo do EM pode ser escrito como

$$\begin{aligned} N \frac{\partial P}{\partial t} &= \frac{\partial \Pi}{\partial t} - \frac{1}{\sqrt{N}} \left\{ \left[N \frac{d\phi_s}{dt} + (b-d)\phi_s \right] \frac{\partial \Pi}{\partial \xi_s} + \right. \\ &\quad \left. + \left[N \frac{d\phi_x}{dt} + (b-d)\phi_x \right] \frac{\partial \Pi}{\partial \xi_x} \right\} + \\ &\quad + \frac{1}{N} \frac{(b-d)}{2} \left[\frac{\partial (\xi_s \Pi)}{\partial \xi_s} + \frac{\partial (\xi_x \Pi)}{\partial \xi_x} \right]. \end{aligned} \quad (4.16)$$

Podemos observar que agora o operador $N \frac{d}{dt}$ substitui o tempo reescalado (t/N) da expansão usual e que N não pode ser inserido no interior da derivada sem penalidades, uma vez que ela também é uma derivada temporal e o nosso modelo prevê que $\frac{dN}{dt} = b-d$, em vez de ser zero, conforme a Equação (4.4).

O lado direito do EM pode ser obtido usando as mesmas probabilidades de transição dadas pelas regras do nosso modelo (veja a seção 4.2) e segue, praticamente, o mesmo procedimento da expansão usual de van Kampen.

Comparando ambos os lados e coletando as potências da ordem de $N^{-1/2}$, podemos mostrar que

$$\begin{aligned} &\left[N \frac{d\phi_s}{dt} + (b-d)\phi_s \right] \frac{\partial \Pi}{\partial \xi_s} + \left[N \frac{d\phi_x}{dt} + (b-d)\phi_x \right] \frac{\partial \Pi}{\partial \xi_x} = \\ &\frac{\partial \Pi}{\partial \xi_s} [b - (2\lambda\phi_x + d)\phi_s] + \frac{\partial \Pi}{\partial \xi_x} [(2\lambda\phi_s - d)\phi_x]. \end{aligned} \quad (4.17)$$

É fácil verificar que essa equação pode ser desacoplada em duas novas, a saber,

$$\begin{cases} N \frac{d\phi_s}{dt} = b(1 - \phi_s) - 2\lambda\phi_s\phi_x \\ N \frac{d\phi_x}{dt} = -b\phi_x + 2\lambda\phi_s\phi_x, \end{cases} \quad (4.18)$$

que são equivalentes às equações de campo médio obtidas anteriormente quando consideramos que $\phi_s + \phi_x = 1$ (veja a Equação (4.6)).

Seguindo para a próxima ordem da aproximação (N^{-1}), obtemos diretamente uma equação de *Fokker-Planck* (FP) multivariáveis. Tendo em vista que estamos interessados especificamente nas flutuações em torno da solução para tempos grandes, podemos

usar os valores assintóticos de ϕ_s e de ϕ_x , (linhas tracejadas na Figura 4.2a). Uma vez que

$$\phi_s(t \gg 1) \simeq \frac{b}{2\lambda} \quad (4.19)$$

e que

$$\phi_x(t \gg 1) \simeq \left(1 - \frac{b}{2\lambda}\right), \quad (4.20)$$

a equação de FP se torna

$$\begin{aligned} N \frac{\partial \Pi}{\partial t} = & - \frac{\partial}{\partial \xi_s} \left\{ \Pi \left[\left(\frac{b-d}{2} - 2\lambda \right) \xi_s - b \xi_x \right] \right\} + \\ & - \frac{\partial}{\partial \xi_x} \left\{ \Pi \left[\left(\frac{b-d}{2} \right) \xi_x + (2\lambda - b) \xi_s \right] \right\} + \\ & + \frac{1}{2} \left[b + \frac{b}{2\lambda} (d + 2\lambda - b) \right] \frac{\partial^2 \Pi}{\partial \xi_s^2} + (2\lambda - b) \left(\frac{b+d}{4\lambda} \right) \frac{\partial^2 \Pi}{\partial \xi_x^2} + \\ & - \frac{b}{4\lambda} (2\lambda - b) \frac{\partial^2 \Pi}{\partial \xi_s \partial \xi_x}. \end{aligned} \quad (4.21)$$

Resolver essa equação não é simples. Uma forma alternativa pode ser construir a função característica de $\Pi(\cdot)$ usando os momentos estatísticos e, em seguida, calcular $\Pi(\cdot)$ por meio de uma transformada inversa de Fourier. Apesar disso, resultados suficientes já podem ser identificados com os primeiros e os segundos momentos da Equação (4.21).

Os primeiros momentos da equação de FP podem ser escritos na forma

$$N \frac{d\vec{\mu}(t)}{dt} = \mathbf{A} \vec{\mu}(t), \quad (4.22)$$

sendo

$$\vec{\mu}(t) = \begin{bmatrix} \langle \xi_s \rangle \\ \langle \xi_x \rangle \end{bmatrix}$$

e

$$\mathbf{A} = \frac{1}{2} \begin{bmatrix} b - d - 4\lambda & -2b \\ 4\lambda - 2b & b - d \end{bmatrix}.$$

Visto que as condições iniciais, determinísticas, são $\vec{\mu}(0) = \vec{0}$, a solução geral do sistema de equações diferenciais deve ser $\vec{\mu}(t) = \vec{0}$. Esse resultado mostra, explicitamente, que as flutuações $\langle \xi_i \rangle$ estão distribuídas simetricamente em torno dos valores

assintóticos das concentrações ϕ_i calculadas analiticamente. Por consequência, as flutuações das soluções de campo médio para as variáveis S e X também apresentam esse comportamento.

Em seguida, devemos investigar os segundos momentos. Analogamente ao momento anterior, eles também satisfazem um sistema de equações diferenciais. A saber,

$$N \frac{d\vec{\sigma}(t)}{dt} = \mathbf{B} \vec{\sigma}(t) + \vec{g}, \quad (4.23)$$

sendo

$$\vec{\sigma}(t) = \begin{bmatrix} \langle \xi_s^2 \rangle \\ \langle \xi_x^2 \rangle \\ \langle \xi_s \xi_x \rangle \end{bmatrix},$$

$$\mathbf{B} = \begin{bmatrix} b - d - 4\lambda & 0 & -2b \\ 0 & b - d & 4\lambda - 2b \\ 2\lambda - b & -b & b - d - 2\lambda \end{bmatrix},$$

e

$$\vec{g} = \frac{1}{4\lambda} \begin{bmatrix} 4\lambda b + 2b(d + 2\lambda - b) \\ 2(2\lambda - b)(b + d) \\ -b(2\lambda - b) \end{bmatrix}.$$

Esse sistema é linear e tem uma solução analítica em termos dos conjuntos $\{u_i\}$ e $\{\vec{v}_i\}$ ($i = 1, 2, 3$), que representam, respectivamente, os autovalores e autovectores de $\vec{B}$. Assim, podemos utilizar a matriz constante T de modo que $\vec{T}^{-1} \vec{B} \vec{T}$ seja diagonal para definir a variável $\vec{y}$, como

$$\vec{\sigma}(t) = T \vec{y}(t). \quad (4.24)$$

Por construção, as colunas de $\vec{T}$ são os próprios autovectores de $\vec{B}$. Desse modo, podemos reescrever a Equação (4.23) na forma

$$N \frac{dy_i}{dt} = u_i(y_i + f_i), \quad (4.25)$$

em que $i = 1, 2$ e 3 e $u_i f_i$ é a i -ésima componente do vetor $T \vec{g}$.

A Equação (4.25) pode ser resolvida, levando à solução

$$|y_i + f_i| = \tilde{c}_i N^{\frac{u_i}{b-a}}, \quad (4.26)$$

em que $\tilde{c}_i$ é uma constante que deverá ser ajustada de acordo com as condições iniciais determinísticas $\langle \xi_i^2 \rangle = 0$. Dessa forma, a Equação (4.23) tem uma solução analítica fechada como uma combinação linear de N elevado a potências proporcionais aos autovalores de B , e seu comportamento segue para um valor aproximadamente constante para tempos longos. A solução numérica dessa equação está ilustrada na Figura 4.5a e pode ser usada para calcularmos as variâncias das concentrações simuladas, que, por sua vez, são dadas por

$$\langle \phi_i^2 \rangle = \frac{\langle \xi_i^2 \rangle}{\langle N \rangle}, \quad (4.27)$$

em que $i = S$ ou X e $\langle N \rangle$ é a população média. A Figura 4.5b ilustra o comportamento de $\langle \phi_s^2 \rangle$ e $\langle \phi_x^2 \rangle$.

Visto que $\langle \xi_i \rangle = 0$ e que $\langle \xi_i^2 \rangle \propto t^{-\alpha}$ com $\alpha \simeq 0.96$ para tempos grandes, as flutuações estocásticas em torno dos limites assintóticos são esperadas como um processo de decaimento rápido e simétrico em torno das concentrações obtidas exatamente através das equações de campo médio (Equações (4.8) e (4.9)).

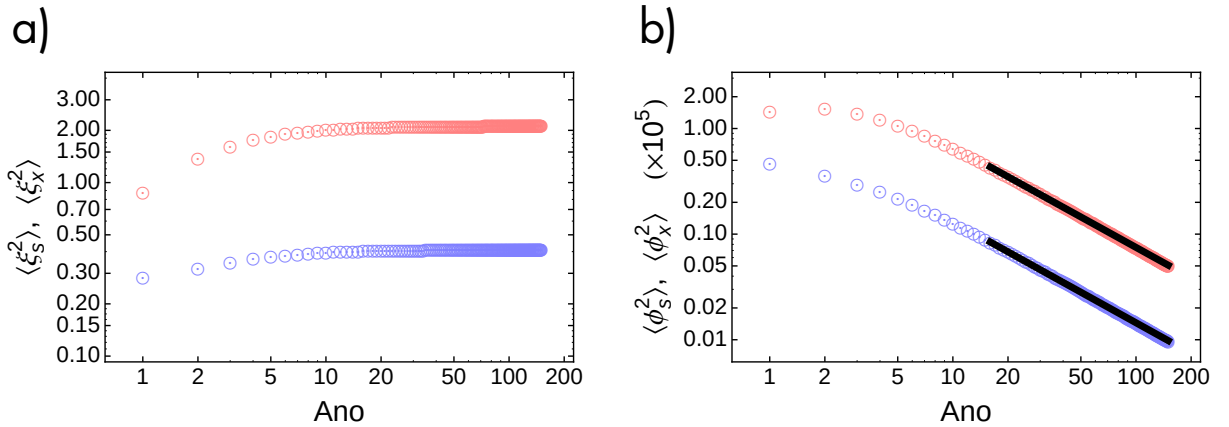


Figura 4.5: Segundos momentos. Resultados obtidos através do modelo para susceptíveis ($\circ$) e para infectados ($\circ$). (a) Segundos momentos das flutuações em torno das concentrações; (b) Segundos momentos das concentrações. As curvas crescem como uma lei de potência negativa para tempos grandes. Um ajuste linear (linhas contínuas) sugere a mesma inclinação ($\simeq 0.96$) para ambas as curvas.

4.5. DISCUSSÃO E PERSPECTIVAS

Nesse capítulo, introduzimos um modelo baseado em uma compartimentalização de agentes (isto é, SI) que considera um fluxo de indivíduos nas classes susceptível (grupo de risco) e infectada (indivíduos contaminados pelo HIV/AIDS). A fim de analisar a dinâmica populacional, introduzimos dois parâmetros: b (representando um fluxo positivo de indivíduos), que está associado principalmente ao crescimento populacional da região infectada e ao ingresso de novos indivíduos na população susceptível; d (representando um fluxo negativo de indivíduos), que, por sua vez, está associado à mudança de comportamento dos indivíduos perante as campanhas em massa contra o HIV/AIDS e à morte dos indivíduos de ambas as classes. Quando ocorre o contato de um agente susceptível com outro infectado, a infecção pode ser transmitida (com uma dada probabilidade λ) e o foco da doença pode crescer e espalhar-se dimensionalmente, apesar de o modelo assumir uma população totalmente misturada que não leva em conta esse fator.

A comparação do resultado do modelo com os dados empíricos está disposta na Figura 4.1b e está em acordo com a Figura 3.3a. Isso indica que os fatores estabilizantes do comportamento logístico se devem, principalmente, à dinâmica populacional gerada tanto naturalmente quanto induzida pelos programas de controle e de combate contra a infecção. Extrapolando o modelo para além dos dados empíricos (Figuras 4.2a e 4.2b), obtemos uma previsão de como poderá ser a dinâmica futura do HIV/AIDS no Brasil.

A Figura 3.3a mostra toda a evolução do HIV no território brasileiro e que, em um cenário geral, as novas taxas de infecção por HIV/AIDS no Brasil estão correndo para um *plateau* (Figura 3.3a). Esse comportamento se deve, primeiramente, à resposta da população ao crescimento da doença. Entretanto, por se tratar de uma doença lenta e não-curável, a dinâmica populacional e as conexões entre indivíduos susceptíveis e infectados têm um papel relevante, conforme ilustramos com os resultados do modelo desse capítulo.

Em particular, também mostramos que um modelo de agentes discretos apresenta uma divergência sistemática do caso contínuo para tempos longos conforme as regras propostas na seção 4.2, mas serve como uma aproximação adequada para os primeiros anos da infecção (*ca.* $\lesssim 12$ anos; Figura 4.3). Por fim, usamos um método formal para analisar

as flutuações estatísticas esperadas para o estado estacionário das solução analítica para as concentrações das espécies, mostrando que elas têm uma distribuição de rápido decaimento (veja a Figura 4.5b) e simétrica em torno do valor esperado para tempos grandes.

DINÂMICA ESPAÇO-TEMPORAL DA DENGUE NO BRASIL

Neste capítulo, iremos abordar a dinâmica espaço-temporal da dengue no Brasil. Tomaremos, como exemplo, alguns métodos usados no capítulo 3 para afrontar diretamente os resultados dessa pandemia.

5.1. A DENGUE NO MUNDO

A dengue é a infecção por vírus mais frequente entre humanos atualmente [106]. O arbovírus¹ da dengue inclui quatro sorotipos imunológicos que causam sintomas similares aos da gripe e, às vezes, complicações que podem ser letais, configurando, nesse último caso, a *dengue hemorrágica* — identificada na década de 1950 [107]. Atualmente, a dengue hemorrágica já afeta a maior parte dos países da Ásia e da América Latina [107]. O principal vetor da doença é o mosquito *Aedes aegypti*² e a sua transmissão entre humanos ocorre por meio de picadas dos mosquitos fêmeas infectados. Humanos infectados são os principais hospedeiros e multiplicadores do vírus, servindo como fonte de contaminação para outros mosquitos, que nascem não-infectados [107]. Uma vez infectado, o mosquito passa por um período de incubação de 4 a 10 dias e irá transmitir a infecção até o resto

¹ *Arbovírus* é um vírus transmitido essencialmente por artrópodes, como os mosquitos.

² Naturalmente, há outras espécies de mosquitos capazes de transmitir o vírus da dengue. Porém, em menor escala, como os *Aedes albopictus* (segundo vetor na Ásia, que, por se adaptar bem a vários climas, já atingiu a Europa e a América do Norte [107]). Vale ressaltar, ainda, que os mosquitos *Aedes aegypti* também podem transmitir outras doenças, como a febre amarela [108].

de sua vida (*ca.* 1,5 – 3 semanas [109]).

5.2. PERFIL DA DENGUE NO MAIOR PAÍS TROPICAL

Sabemos que a dengue é uma doença favorecida fortemente pelos climas tropical e subtropical e pelo crescimento populacional nessas regiões, onde atinge milhões de pessoas [110–112]. Seu maior foco de incidência são as áreas urbanas ou semiurbanas e afeta tanto regiões pobres com baixo índice de desenvolvimento quanto bairros nobres. Atualmente, a forma de enfrentamento à dengue mais eficiente tem sido combater seu vetor.

A incidência global da dengue aumentou cerca de 30 vezes ao longo dos últimos 50 anos [113]. E estima-se que entre 50 e 100 milhões de novas infecções ocorram anualmente nos mais de 100 países endêmicos, colocando mais de 40% da população do globo (*ca.* 2,5 bilhões de pessoas) em risco de contrair a infecção [113]. Já no Brasil, a maior incidência da última década ocorreu em 2009, com um pico de quase 10^6 de novas infecções, isto é, cerca de 0,5% da população brasileira foi infectada pela dengue somente naquele mesmo ano.

5.3. DADOS

Os dados de novas infecções pela dengue que analisaremos nesta seção também foram obtidos diretamente do DATASUS [114]. Especificamente, analisaremos as taxas de incidência anual z , definidas, formalmente, como o número de novos casos de dengue diagnosticados em um determinado ano e cidade. Ao todo, analisaremos dados de quase 6 milhões de casos de dengue no intervalo de 2001 a 2012 para uma soma de 5085 cidades brasileiras com pelo menos um caso de infecção pelo vírus no período de reporte dos dados.

5.4. EVOLUÇÃO TEMPORAL: UM PADRÃO SAZONAL

Diferentemente do caso do HIV/AIDS, cujos dados reportam desde o primeiro caso identificado, os dados da dengue no Brasil representam o comportamento da doença em apenas um período intermediário. Desse modo, não se espera encontrar uma curva de crescimento típica, como a curva logística da Figura 3.3a. A exemplo do que fizemos na seção 3.4, consideramos Z a taxa de incidência total para o t -ésimo ano, obtida da soma das taxas z das $M = 5085$ cidades brasileiras com pelo menos um caso da doença identificado. Assim,

$$Z = \sum_{i=1}^M z_i. \quad (5.1)$$

A Figura 5.1a mostra a evolução mensal de Z no período de reporte de dados. É notável a presença de uma *sazonalidade* com período de 12 meses. Entretanto, os picos nitidamente variam de ano a ano, assim como a sua largura. Tomemos, por exemplo, as taxas de incidência mensais nos anos de 2002 e 2010 (região envolvendo os dois picos mais altos da Figura 5.1a). Ambos os picos têm aproximadamente a mesma amplitude, mas, quando olhamos para a evolução anual dos casos de dengue (Z_{anual} ;³ Figura 5.1b), percebemos que há uma grande diferença entre eles. Como regra, os picos ocorrem anualmente no final do verão ou no início do outono (março e abril), enquanto que os mínimos ocorrem no final do inverno ou no início da primavera (setembro e outubro).

Sabemos que a característica típica de sazonalidade está associada principalmente a variações de fatores ambientais e/ou ecológicos que mudam no ciclo de um ano, como a umidade, o nível de precipitação, a temperatura, as relações parasita-hospedeiro e a dinâmica dos vetores *Aedes* [112, 115, 116]. De um ponto de vista mais formal, a sazonalidade pode ser vista como uma força dirigente, conduzindo a dinâmica de sistemas naturais e as suas populações, o que é essencialmente verdadeiro para o caso de infecções transmitidas por vetores como o *Aedes* [117]. Obviamente, esses aspectos podem mascarar um mecanismo adjacente ao sistema [117, 118]. Para o caso da dengue, estudos mostram que a sazonalidade na incidência da dengue pode ser associada à transição de

³ A definição do número de casos anual é trivialmente definida como $Z_{anual} = \sum_{j=\text{janeiro}}^{\text{dezembro}} z_j$

um parâmetro $\lambda(t)$, que assume diferentes valores durante as fases epidêmica ($\lambda > \lambda_c$) e endêmica ($\lambda < \lambda_c$) [116].

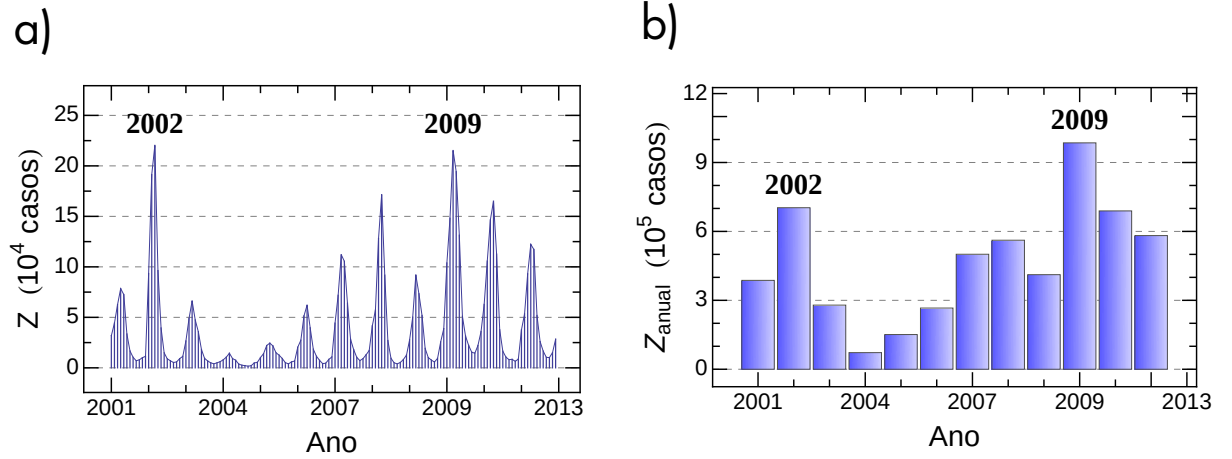


Figura 5.1: Evolução temporal dos casos de dengue no Brasil. (a) Número mensal de novos casos de dengue identificados no período de janeiro de 2001 a dezembro de 2012. (b) Número anual de novos casos de dengue identificados no mesmo período.

5.5. DISTRIBUIÇÃO DAS TAXAS DE INCIDÊNCIA

Nesta seção, faremos uma investigação sobre a correlação do vírus da dengue com o tamanho dos centros urbanos infectados pela doença. A Figura 5.2 mostra a distribuição dos casos de dengue no Brasil no ano de 2010 normalizada pela densidade populacional das cidades nesse mesmo ano, segundo o censo do IBGE [81]. Para visualizarmos melhor a comparação, a distribuição de cores está em escala logarítmica e, nos eixos, há distribuição da incidência de dengue por latitude (horizontal) e por longitude (vertical). A parte mais atingida é o extremo Nordeste, mas uma densidade considerável atingiu a região Centro-oeste (longitude $\sim 51^\circ$) e a Sudeste (latitude $\sim -21^\circ$).

Com o objetivo de compreender melhor a incidência da dengue sobre as cidades, usamos o mesmo procedimento feito com o HIV/AIDS: analisamos a função densidade de probabilidade das taxas de incidência da dengue entre cidades brasileiras $P(z)$. A Figura 5.3a mostra as distribuições de probabilidade $P(z)$ em comparação com um decaí-

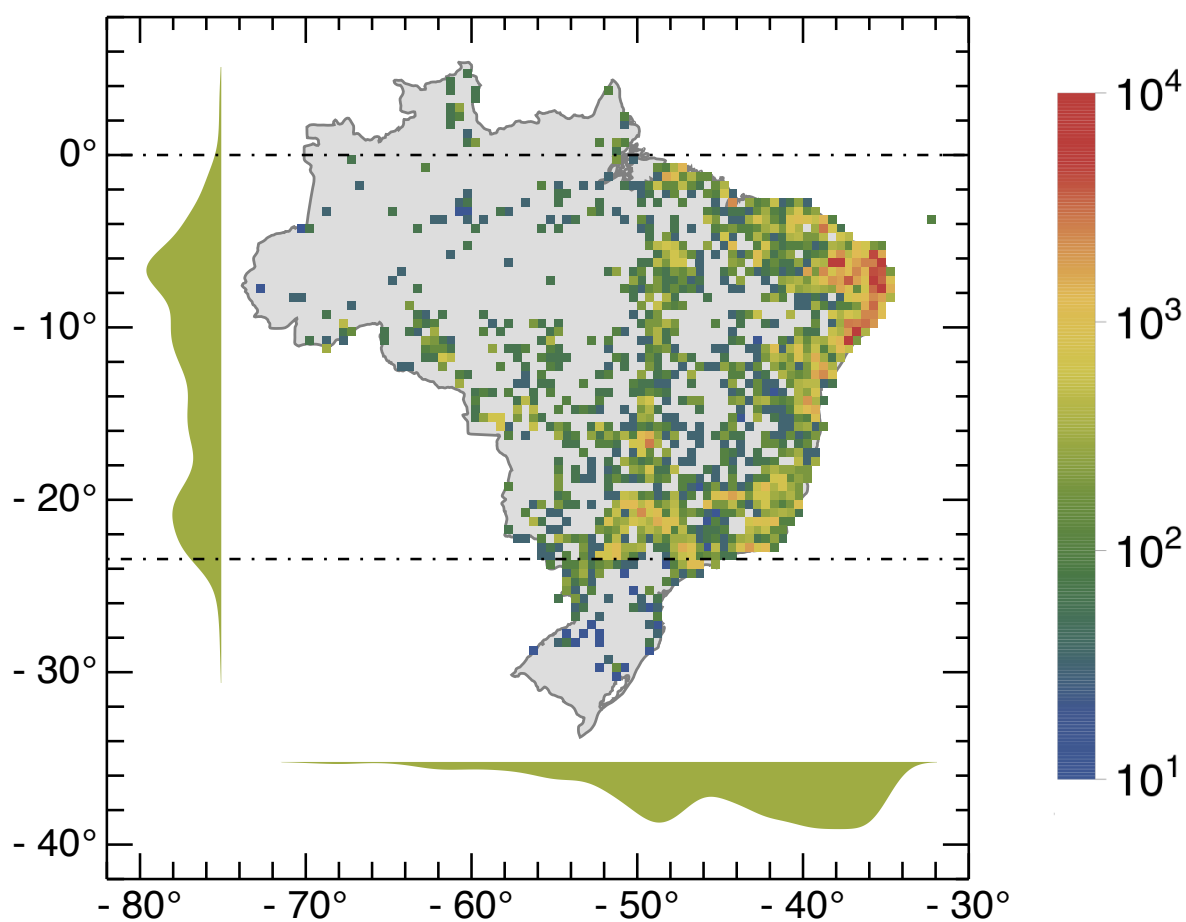


Figura 5.2: Densidade da dengue pelo número de habitantes. Incidência do vírus da dengue em 2010 normalizada pela população das cidades ($\times 10^6$). Nos eixos, há uma distribuição da incidência de dengue por latitude (horizontal) e por longitude (vertical). As linhas tracejadas correspondem à linha do Equador e a do Trópico de Capricórnio.

mento tipo lei de potência na forma

$$P(z) \propto z^{-\alpha}. \quad (5.2)$$

Na Figura 5.3a, percebemos que as distribuições têm uma cauda do tipo lei de potência com $\alpha \sim 1.87$, que permanece robusta por cerca de três décadas. Os expoentes α para cada ano permanecem sistematicamente constantes durante todo o período de reporte de dados, conforme ilustrado na Figura 5.3b.

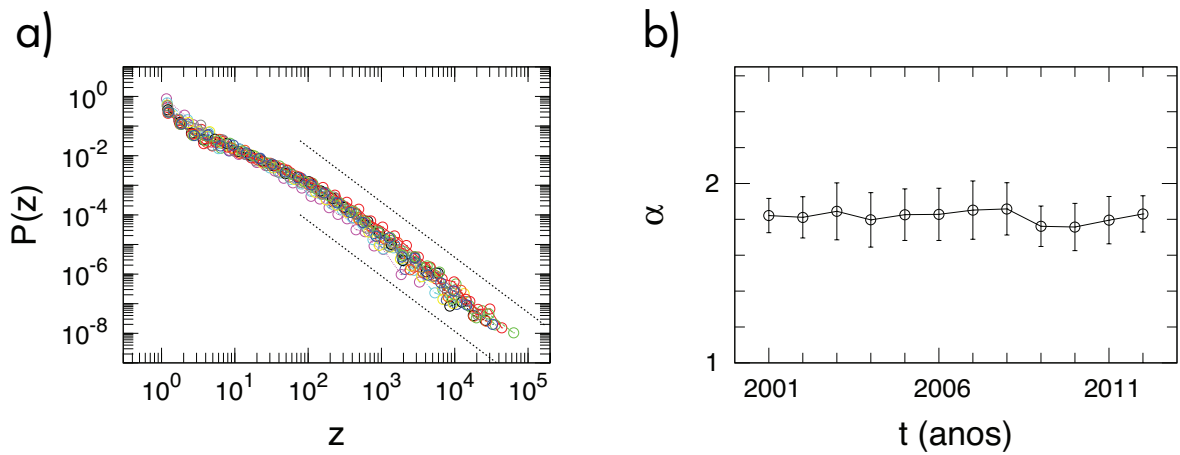


Figura 5.3: PDF das taxas de incidência nas cidades. (a) Gráfico em escala log-log das funções densidade de probabilidade $P(z)$, em que z corresponde às taxas de incidência da infecção e cada cor representa um ano. As linhas tracejadas têm inclinação -1.87 e servem de guia para a cauda do tipo lei de potência. (b) Evolução temporal do expoente α das leis de potência para o período de 2001 a 2012, com barras de erro para 99% de confiança.

5.6. TAXAS DE INCIDÊNCIA VERSUS POPULAÇÃO

Conforme ilustrado na Figura 5.2, o número total de indivíduos infectados *per capita* depende de diferentes fatores, como a concentração dos vetores e as variações climáticas, e não deve ser constante em diferentes cidades. Na seção 3.6, encontramos uma relação alométrica, Equação (3.6), entre as taxas de incidência do HIV/ADIS e a população das cidades. Novamente, usando a distribuição da população brasileira fornecida pelo IBGE [81], obtivemos a Figura 5.4, que mostra as taxas de incidência z da dengue em cada cidade como uma função do seu número de habitantes p . Para uma melhor aproxi-

mação, selecionamos somente as cidades com $p \geq 3 \times 10^3$. A curva obtida é novamente bem descrita por uma lei de escala na forma

$$z \propto p^\delta. \quad (5.3)$$

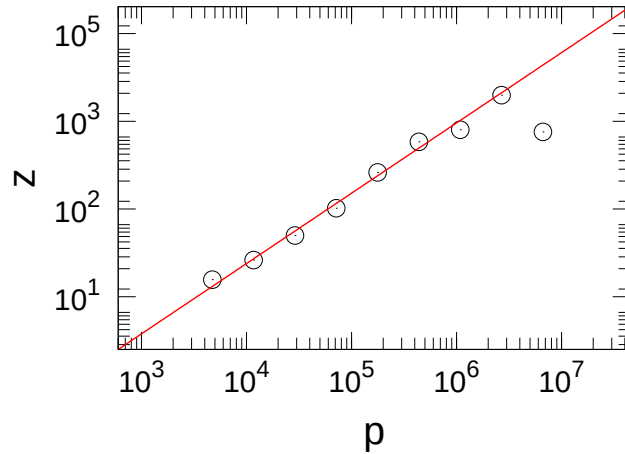


Figura 5.4: Relação alométrica entre as taxas de incidência e a população. Taxas de incidência z (usando janelas log-espaciações) da dengue como uma função da população p das cidades mostradas em escala log-log para o ano 2010 nas cidades com $p \geq 3 \times 10^3$. A linha contínua é um ajuste linear com inclinação 1.01 ± 0.03 .

Particularmente, encontramos $\delta \simeq 1$, ou seja, a dengue escala linearmente com a população da região infectada, diferentemente do que acontece com o HIV/AIDS que apresenta uma dependência superlinear. Isso deve estar relacionado com o fato de o mosquito picar os hospedeiros diferentemente de seu *status* social ou comportamento. Naturalmente, há uma notável distinção no comportamento do grupo de risco do HIV, o que parece ser desprezível para o vírus da dengue.

5.7. DISCUSSÃO E PERSPECTIVAS

Notadamente, a Figura 3.3a mostra a evolução do HIV no território brasileiro desde o primeiro caso. Em contraste, a Figura 5.1 mostra um corte na evolução da dengue no país em apenas um período intermediário. É notável a presença do padrão sazonal nas taxas de incidência da dengue. Isso se deve, por exemplo, às variações climáticas ao longo de um ano, como o nível de precipitação e de umidade, as variações de tem-

peratura e a dinâmica populacional dos vetores *Aedes*, bem como a sua relação com os hospedeiros [112, 115, 116, 119].

Também mostramos, nesse capítulo, que, em pequenas escalas espaciais (em nível de cidades), a distribuição da infecção pela dengue tem um comportamento do tipo lei de potência robusta para todo o período com dados com expoente negativo médio $\alpha \simeq 1,87$, conforme ilustrado na Figura 5.3. Esse comportamento é bastante similar ao encontrado para o HIV/AIDS no Capítulo 3, no qual identificamos, aproximadamente, o mesmo expoente negativo: $\alpha \simeq 1,9$. O surgimento de comportamentos do tipo lei de potência é um padrão distinto: leis de potência decaem mais lentamente do que regimes exponenciais e são características ubíquas da natureza. Vale notar, ainda, que os expoentes estão sistematicamente abaixo da lei do inverso do quadrado ($P(y) \propto y^{-2}$), uma característica onipresente na natureza.

Com o objetivo de comparar a distribuição das taxas de incidência da dengue no território brasileiro, analisamos como elas escalam com a população da região onde estão centradas. Na Figura 5.2, as taxas de incidência da dengue estão normalizadas pela população da cidade, mostrando que a distribuição de casos não é fortemente dependente da densidade populacional. Também, percebemos, pelos histogramas nessa figura, que há grande incidência da dengue no extremo Nordeste e no Sudeste. Em relação à população das cidades, preparamos a Figura 5.4, que ilustra a regra alométrica entre a taxa de incidência da dengue contra a população das cidades, por meio de médias em janelas log-espaçadas na população. Para o caso da dengue, a lei de escala é linear $\delta \simeq 1$, assim como as taxas de mortalidade por *influenza* e por pneumonia em cidades dos Estados Unidos em torno de 1918 [120]. Em contraste, mostramos, no capítulo 3, que, para o caso do HIV/AIDS, essa lei de escala mostrou-se superlinear com expoente $\delta \simeq 4/3$ (veja a Figura 3.5) [83, 84].

CONCLUSÃO

De maneira geral, nesta tese empregamos dados empíricos com o objetivo de identificar padrões e de desenvolver modelos matemáticos. A abordagem específica de cada problema foi projetada de acordo com o perfil dos dados disponíveis. No caso do acidente do Golfo do México, por exemplo, havia apenas uma série temporal para cada grupo de animais, o que limitou as possibilidades de análise. Entretanto, a identificação do crescimento exponencial das taxas de mortalidade indicaram, quantitativamente, que vazamentos prolongados de petróleo podem ter um efeito devastador sobre a biota marinha e costeira.

Os dados das loterias também apresentaram algumas limitações, mas, graças às regras de distribuição de prêmios serem bem definidas, foi possível elaborar a Figura (2.2) e identificar um padrão linear nela, que foi o resultado mais importante do capítulo 2, visto que ele nos fornece um mecanismo gerador de correlações para as séries de ganhadores. Obviamente, essa presença de persistência em um sistema supostamente aleatório é um assunto intrigante, ainda mais quando recursos financeiros estão envolvidos.

Em contraste com os dados do acidente no Golfo do México e de loterias, os dados referentes às epidemias foram muito mais fartos, o que permitiu mais formas de ataque aos problemas. Por exemplo, foi possível construir distribuições de ocorrência em nível de cidades e checar as leis de escala de doenças segundo a densidade populacional da região. Foi o caso do crescimento logístico do HIV no Brasil, que é justificável por meio de um fluxo de indivíduos, conforme identificamos no capítulo 4.

Em resumo, nessa tese, encontramos diversos comportamentos universais, como o crescimento exponencial na taxa de mortalidade de animais no Golfo do México devido ao vazamento de petróleo com taxa aproximadamente constante na região. Outro crescimento exponencial foi encontrado no Capítulo 2: o aumento das vendas depois de subsequentes acumulações da *sena*. Usamos, nesse caso, um modelo estocástico para mostrar que esse crescimento é um mecanismo gerador de correlações. Podemos espe-

rar, ainda, que mecanismos similares emergjam em outros sistemas, como em um processo de caça, no qual o desejo/necessidade pela/da comida faz o esforço da caça ser maior (enquanto o animal tiver forças suficientes para tanto). Já nos capítulos 3 e 5, encontramos comportamentos robustos para as incidências do HIV/AIDS e da dengue no Brasil, como a presença singular de comportamentos do tipo lei de potência para distribuições em cidades. Mostramos, ainda, que as doenças não estão distribuídas uniformemente sobre o país, que o tamanho das cidades pode influenciar no número de novas infecções e que a lei de escala alométrica dessas grandezas pode não ser a mesma para diferentes doenças. No nosso estudo, verificamos, em particular, que o comportamento superlinear do HIV/AIDS sugere que o vírus depende de grupos sociais e da interação de seus membros, enquanto que a dengue (que é transmitida por um vetor) não é sensível a tal propriedade. No capítulo 4, por sua vez, elaboramos um modelo estocástico de agentes que mostra que a dinâmica populacional influencia fortemente o número de novas infecções pelo HIV.

Por fim, esperamos que os resultados oriundos dessa pesquisa possam ser úteis, por exemplo, na compreensão da dinâmica espaço-temporal de epidemias em geral e na escolha de estratégias adequadas de controle. Ainda, além de padrões e de comportamentos coletivos, problemas envolvendo loterias tangem à forma na qual incentivos influenciam as tomadas de decisões, um problema ainda em aberto em neurociência. Nesse contexto, a Figura (2.2) também representa uma medida quantitativa de como o dinheiro altera o comportamento humano. Por outro lado, esse resultado parece sugerir que o mecanismo de crescimento nas vendas de bilhetes de loteria pode ser associado, pelo menos no contexto de loterias, a uma medida indireta da ambição.

CORRELAÇÕES [121]

Quando variáveis estocásticas são independentes, a estacionariedade¹ implica que o processo estocástico $x(t)$ ($t \in \mathbb{R}$) seja independente do tempo e identicamente distribuído. Grandezas observáveis caracterizando tal processo podem ser escritas em termos dos momentos de n -ésima ordem. O caso $n = 1$ é suficiente para definir a média

$$\langle x(t) \rangle \equiv \int_{-\infty}^{\infty} x f(x, t) dx, \quad (\text{A.1})$$

em que $f(x, t)$ representa a densidade de probabilidade de se observar uma variável aleatória de valor x no tempo t .

O caso $n = 2$ é usado para definir $R(\tau)$, a *função de correlação*²

$$R(\tau) \equiv \langle x(t_1) x(t_2) \rangle \equiv \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f(x_1, x_2; t_1, t_2) dx_1 dx_2, \quad (\text{A.2})$$

em que $f(x_1, x_2; t_1, t_2)$ é a densidade de probabilidade de se observar x_1 no tempo t_1 e x_2 no tempo t_2 com $\tau = t_2 - t_1$. Para caracterizar de maneira plena as propriedades estatísticas de um processo estocástico, devemos conhecer a função $f(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n)$ para todos os x_i, t_i e n . A maior parte dos estudos é limitada a considerar apenas até $n = 2$.

Apesar de apresentar flutuações aleatórias, uma série temporal (sinal ou ruído, por exemplo) pode conter algum tipo de correlação e apresentar diferentes propriedades estatísticas. Visto que a função de correlação relaciona o processo estocástico em dois

¹ Estritamente falando, um processo é dito estacionário se sua PDF for invariante perante uma translação temporal. Desse modo, sua média e variância não devem depender do tempo.

² Alguns autores reescalam a função de correlação como $R(\tau) \rightarrow \frac{R(\tau)}{R(0)}$.

tempos separados por uma distância τ , é comum associar correlações com um efeito de memória. Ou seja, um sistema correlacionado tem lembrança de seu passado e isso se relaciona com o estado futuro de algum modo. Para variáveis estocásticas positivamente correlacionadas, $R(\tau)$ é tipicamente uma função decrescente. E a integral dessa função sobre todos os intervalos de tempo τ pode ser usada para caracterizar a correlação do sistema. Quando essa integral é finita, existe um tempo de memória típico, o *tempo de correlação*³ do processo $x(t)$. E as correlações são ditas *correlações de curto alcance*⁴. Um caso típico é quando

$$R(\tau) = \exp\left(-\frac{\tau^\nu}{\tau_c}\right), \quad (\text{A.3})$$

com $\nu > 0$. O que conduz a

$$\int_0^\infty R(\tau) d\tau \propto \tau_c^{\frac{1}{\nu}}. \quad (\text{A.4})$$

Já quando a integral de $R(\tau)$ é infinita, não existe um tempo de memória típico e as correlações são ditas *correlações de longo alcance*⁵. Um caso típico é quando

$$R(\tau) \propto \tau^{\eta-1}, \quad (\text{A.5})$$

sendo $0 < \eta \leq 1$. O caso particular em que $R(\tau)$ é uma constante independente de τ está ligado a um processo totalmente aleatório que não tem lembrança do passado.

Para o caso de uma série temporal $\{x_i\}$ ($i \in \mathbb{N}$), a função de correlação⁶ poderá ser escrita como

$$r_k = \frac{c_k}{c_0}, \quad (\text{A.6})$$

em que

$$c_k = \frac{1}{N-k-1} \sum_{i=1}^{N-k} (x_i - \langle x_i \rangle) (x_{i+k} - \langle x_i \rangle) \quad (\text{A.7})$$

é a *covariância* de $\{x_i\}$. Podemos notar que c_0 é a própria *variância* de $\{x_i\}$.

³ Assumimos a tradução *comprimento de correlação* para o caso espacial.

⁴ Do inglês *short-range correlations*.

⁵ Do inglês *long-range correlations*.

⁶ No caso discreto, também é comum chamá-la de *coeficiente de correlação*.

A.1. DETRENDED FLUCTUATIONS ANALYSIS – DFA

Muitas vezes, analisar o efeito de memória presente na dinâmica temporal de uma série temporal $\{x_i\}$ ($i \in \mathbb{N}$) por meio da definição de $R(\tau)$ não é trivial, visto que diferentes tendências podem estar presentes na série e até mesmo induzir a falsas conclusões. Desse modo, é muito comum usar métodos alternativos como o *Detrended Fluctuation Analysis* (DFA) em vez da Equação (A.6) [37]. Esse método foi cunhado para remover tendências da série temporal e tem sido amplamente aplicado na literatura para investigar o expoente de Hurst (h), até mesmo em séries temporais não-estacionárias [37, 118]. Basicamente, o DFA- l consiste em quatro etapas:

- (i) reduzir as flutuações acumulando a série temporal: $y_i = \sum_{j=1}^i x_j$;
- (ii) dividir y_i em janelas de tamanho s ;
- (iii) remover tendências locais não-estacionárias, sob a forma de uma função polinomial de grau l para cada janela;
- (iv) calcular o desvio padrão médio sobre todas as janelas de tamanho s , sendo denotado por $F(s)$.

Tipicamente, séries temporais que exibem propriedades fractais são caracterizadas por um comportamento de lei de potência na forma

$$F(s) \propto s^h, \quad (\text{A.8})$$

em que h é um expoente de escala que normalmente indica o valor do expoente de Hurst. O valor numérico de h quantifica presença de autocorrelação da série temporal. Ela terá correlações de longo alcance se $h \neq 0,5$, sendo persistente (antipersistente) quando $h > 0,5$ ($0 < h < 0,5$). Caso contrário, se $h = 0,5$, a série temporal é sem memória ou tem apenas correlação de curto alcance. Em geral, para confirmar a presença de autocorrelação, é comum aplicar diversas vezes o DFA sobre a série original embaralhada. Em um embaralhamento aleatório, a correlação presente deve ser destruída e o expoente de Hurst obtido deve ser próximo de 0,5.

SÉRIES HIPERGEOMÉTRICAS [122]

Em geral, as funções hipergeométricas podem ser escritas como

$${}_pF_q \left[\begin{matrix} a_1 & \cdots & a_p \\ b_1 & \cdots & b_q \end{matrix} ; z \right] = \sum_{k=0}^{\infty} \frac{(a_1)_k \cdots (a_p)_k}{(b_1)_k \cdots (b_q)_k} \frac{z^k}{k!}, \quad (\text{B.1})$$

em que $\Gamma(n)$ representa a função gama e $(\alpha)_k$ é o símbolo de Pochhammer, definido como

$$(\alpha)_k = \frac{\Gamma(\alpha + k)}{\Gamma(\alpha)}, \quad (\text{B.2})$$

sendo $\alpha \in \mathbb{C} \setminus \mathbb{Z}_{\leq 0}$ e $k \in \mathbb{N}$.

B.1. SÉRIE HIPERGEOMÉTRICA DE GAUSS

Partindo da definição (B.1), a Série Hipergeométrica de Gauss é escrita como

$${}_2F_1 \left[\begin{matrix} \alpha & \beta \\ \gamma \end{matrix} ; z \right] = \sum_{n=0}^{\infty} \frac{(\alpha)_n (\beta)_n}{(\gamma)_n} \frac{z^n}{n!}, \quad (\text{B.3})$$

sendo $\{\alpha \vee \beta \vee \gamma\} \in \mathbb{C} \setminus \mathbb{Z}_{\leq 0}$, $n \in \mathbb{N}$ e $z \in \mathbb{C}$. Em particular, ela é a solução da equação diferencial

$$\frac{d^2 w}{dz^2} + \frac{c - (a + b + 1)z}{z(1 - z)} \frac{dw}{dz} - \frac{ab}{z(1 - z)} w = 0. \quad (\text{B.4})$$

Além disso, a Equação (B.3) pode ser escrita em uma forma integral do tipo

$${}_2F_1 \left[\begin{matrix} \alpha & \beta \\ \gamma \end{matrix} ; z \right] = \frac{\Gamma(\gamma)}{\Gamma(\beta)\Gamma(\gamma-\beta)} \int_0^1 t^{\beta-1} (1-t)^{\gamma-\beta-1} (1-zt)^{-\alpha} dt \quad (\text{B.5})$$

para $\{z \in \mathbb{C} : |z| < 1\}$.

TESTE CHI-QUADRADO DE PEARSON [123]

O *Teste Chi-Quadrado* (χ^2) de Pearson é um dos mais conhecidos entre os vários testes estatísticos, e a adequação entre duas curvas é avaliada por referência à distribuição chi-quadrado. Suas propriedades foram primeiramente investigadas por Karl Pearson em torno de 1900.

O teste chi-quadrado de Pearson se resume a formular uma hipótese nula H_0 e checar se há evidência estatística para sua rejeição. O procedimento é feito comparando os dados obtidos experimentalmente (ou simulados) com os dados esperados de acordo com uma lei determinística na forma

$$X^2 = \sum_{i=1}^N \frac{(O_i - E_i)^2}{E_i}, \quad (\text{C.1})$$

em que:

- ➡ $\{O_i\}$ é o conjunto dos dados observados;
- ➡ $\{E_i\}$ é o conjunto esperado;
- ➡ $N \in \mathbb{N}$ é a quantidade de dados em uma amostra.

Em seguida, compara-se o X^2 com um valor crítico χ_α^2 obtido diretamente da distribuição χ^2 , considerando um nível de significância α e $\nu = N - 1$ graus de liberdade do sistema.¹ Nesse cenário, a hipótese nula será descartada se, e somente se,

$$X^2 \geq \chi_\alpha^2. \quad (\text{C.2})$$

¹ Sendo o grau de liberdade ν o parâmetro da distribuição χ .

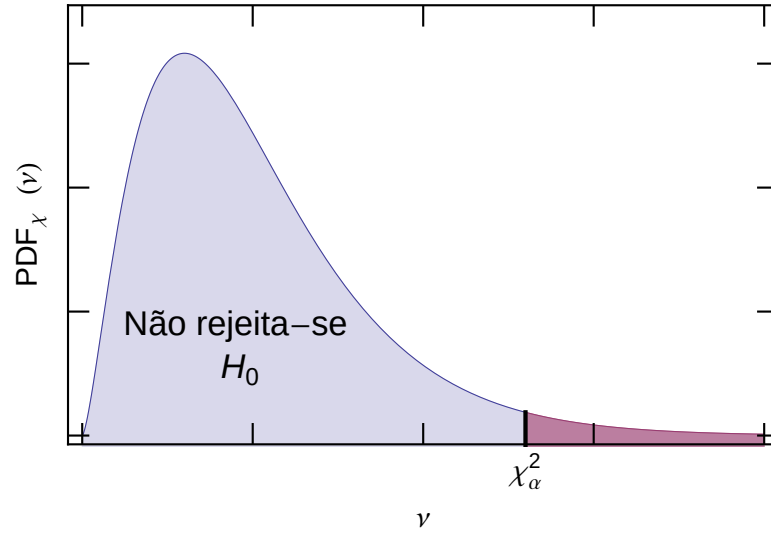


Figura C.1: Distribuição χ^2 com ν graus de liberdade. O ponto crítico χ_α^2 delimita a região de veracidade da hipótese nula (região sombreada em azul).

Graficamente, a condição (C.2) é apresentada na Figura C.1.

A fim de simplificar o processo, é comum os *software* matemáticos informarem diretamente os valores- p , que são calculados como segue:

$$\text{valor-}p \equiv P(\chi^2 > X^2) = \int_{X^2}^{\infty} \text{PDF}_\chi(\nu) dx, \quad (\text{C.3})$$

com $\text{PDF}_\chi(\nu) = x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}}$ ($x \in \mathbb{R}_{\geq 0}$) sendo a distribuição chi-quadrado com grau de liberdade ν . Podemos notar que essa expressão representa uma distribuição acumulada de probabilidade (CDF), portanto, os valores- p estão sempre no intervalo $[0, 1]$.

Assim, podemos compará-los diretamente com a significância estatística. A saber, a hipótese nula será descartada se, e somente se,

$$\text{valor-}p > 1 - \alpha. \quad (\text{C.4})$$

Em geral, temos que $\alpha = 0,01$, com limite máximo de $\alpha = 0,05$.

REFERÊNCIAS

- [1] NEWMAN, M. E. J.; GIRVAN, M. Finding and evaluating community structure in networks. *Phys. Rev. E*, v. 69, n. 2, p. e026113, 2004.
- [2] CASTELLANO, C.; FORTUNATO, S.; LORETO, V. Statistical physics of social dynamics. *Rev. Mod. Phys.*, v. 81, n. 2, p. 591 – 646, 2009.
- [3] CORONEL-BRIZIO, H. et al. Statistical auditing and randomness test of lotto-type games. *Physica A*, v. 387, n. 25, p. 6385 – 6390, 2008.
- [4] ARIYABUDDHIPHONGS, V. Lottery gambling: A review. *J. Gambl. Stud.*, v. 27, n. 1, p. 15 – 33, 2011.
- [5] HSU, M. et al. Neural systems responding to degrees of uncertainty in human decision-making. *Science*, v. 310, p. 1680 – 1683, 2005.
- [6] BRIERS, B. et al. Hungry for money: the desire for caloric resources increases the desire for financial resources and vice versa. *Psychol. Sci.*, v. 17, p. 939 – 943, 2006.
- [7] PESSIGLIONE, M. et al. How the brain translates money into force: A neuroimaging study of subliminal motivation. *Science*, v. 316, n. 5826, p. 904 – 906, 2007.
- [8] HALL-STOODLEY, L.; COSTERTON, J. W.; STOODLEY, P. Bacterial biofilms: from the Natural environment to infectious diseases. *Nat. Rev. Micro.*, v. 2, n. 2, p. 95 – 108, 2004.
- [9] PALESE, P. Influenza: old and new threats. *Nat. Med.*, v. 10, n. 12, p. S82 – S87, 2004.
- [10] ANTONIO, F. J.; MENDES, R. S.; THOMAZ, S. M. Identifying and modeling patterns of tetrapod vertebrate mortality rates in the Gulf of Mexico oil spill. *Aquat. Toxicol.*, v. 105, n. 1–2, p. 177 – 179, 2011.

- [11] BOHANNON, J.; BOSCH, X.; WITHGOTT, J. Scientists brace for bad tidings after spill. *Science*, v. 298, n. 5599, p. 1695 – 1695, 2002.
- [12] CASTEGE, I. et al. Estimating actual seabirds mortality at sea and relationship with oil spills: lesson from the “Prestige” oil spill in aquitaine (France). *Ardeola*, v. 54, n. 2, p. 289–307, 2007.
- [13] SAMANTA, S. K.; SINGH, O. V.; JAIN, R. K. Polycyclic aromatic hydrocarbons: environmental pollution and bioremediation. *Trends Biotechnol.*, v. 20, n. 6, p. 243 – 248, 2002.
- [14] BILBAO, E. et al. Effects of exposure to Prestige-like heavy fuel oil and to perfluorooctane sulfonate on conventional biomarkers and target gene transcription in the thicklip grey mullet *Chelon labrosus*. *Aquat. Toxicol.*, v. 98, n. 3, p. 282 – 296, 2010.
- [15] USFWS. Disponível em: <www.fws.gov/home/dhoilspill/collectionreports.html>. Acesso em: dezembro de 2010.
- [16] The Telegraph. Disponível em: <<http://www.telegraph.co.uk/earth/wildlife/7713466/Gulf-of-Mexico-oil-spill-dead-dolphins-found-washed-up-on-US-coast.html>>. Acesso em: maio de 2010.
- [17] NOAA – STSSN. Disponível em: <<http://www.sefsc.noaa.gov/stssnrep/SeaTurtleReportI.do?action=reportquery>>. Acesso em: janeiro de 2011.
- [18] NMFS. Disponível em: <http://www.nmfs.noaa.gov/pr/health/mmume/cetacean_gulfofmexico2010.htm>. Acesso em: maio de 2011.
- [19] MURRAY, J. D. *Mathematical Biology: I. An Introduction*. 3. ed. [S.l.]: Springer, 2002. ISBN 0387952233.
- [20] History Of Lottery. Disponível em: <<http://historyoflottery.com/index.html>>. Acesso em: agosto de 2013.
- [21] Learn about lotto history | Lotto Madness Slots. Disponível em: <<http://www.lottomadnessslots.com/lotto-madness-history.html>>. Acesso em: agosto de 2013.

- [22] FARRELL, L.; MORGENROTH, E. L. W.; WALKER, I. A time series analysis of UK lottery sales: the long run price elasticity. *Oxford B. Econ. Stat.*, v. 61, p. 513 – 526, 1999.
- [23] Mega-Sena – Resultado. Disponível em: <http://www1.caixa.gov.br/loterias/loterias/megasena/megasena_resultado.asp>. Acesso em: abril de 2013.
- [24] MALTBY, J. et al. Implicit theories of a desire for fame. *Brit. J. Psychol.*, v. 99, n. 2, p. 279 – 292, 2008.
- [25] SIMON, B.; OAKES, P. Beyond dependence: an identity approach to social power and domination. *Hum. Relat.*, v. 59, p. 105 – 139, 2006.
- [26] SANFEY, A. G. Social decision-making: Insights from game theory and neuroscience. *Science*, v. 318, n. 5850, p. 598 – 602, 2007.
- [27] LEA, S. E. G.; WEBLEY, P. Money as tool, money as drug: The biological psychology of a strong incentive. *Behav. Brain Sci.*, v. 29, p. 161 – 209, 2006.
- [28] FEHR, E.; FALK, A. Psychological foundations of incentives. *Eur. Econ. Rev.*, v. 46, 2002.
- [29] BÉNABOU, R.; TIROLE, J. Intrinsic and extrinsic motivation. *Rev. Econ. Stud.*, v. 70, p. 489 – 520, 2003.
- [30] VOHS, K. D.; MEAD, N. L.; GOODE, M. R. The psychological consequences of money. *Science*, v. 314, n. 5802, p. 1154 – 1156, 2006.
- [31] CAIXA – Loterias – Como Jogar. Disponível em: <http://www1.caixa.gov.br/loterias/loterias/megasena/como_jogar.asp>. Acesso em: abril de 2013.
- [32] BAKER, R. D.; MCHALE, I. G. Modelling the probability distribution of prize winnings in the UK national lottery: consequences of conscious selection. *J. R. Statist. Soc. A.*, v. 172, n. 4, p. 813 – 834, 2009.
- [33] CAIXA – Loterias – Distribuição de arrecadação. Disponível em: <<http://www1.caixa.gov.br/loterias/loterias/megasena/distribuicao.asp>>. Acesso em: abril de 2013.

- [34] CAIXA – Loterias – Mega da Virada. Disponível em: <http://www1.caixa.gov.br/loterias/loterias/megasena/mega_virada.asp>. Acesso em: abril de 2013.
- [35] MCHALE, I.; PEEL, D. Habit and long memory in UK lottery sales. *Econ. Lett.*, v. 109, n. 1, p. 7 – 10, 2010.
- [36] DEBIT. Disponível em: <<http://www.debit.com.br/consulta30.php?acumulado=1&indice=inpc>>. Acesso em: abril de 2013.
- [37] PENG, C.-K. et al. Long-range correlations in nucleotide sequences. *Nature*, v. 356, n. 6365, p. 168 – 170, 1992.
- [38] TOM, S. M. et al. The neural basis of loss aversion in decision-making under risk. *Science*, v. 315, n. 5811, p. 515–518, 2007.
- [39] DOYA, K. Modulators of decision making. *Nat. Neurosci.*, v. 11, n. 4, p. 410 – 416, 2008.
- [40] LOEWENSTEIN, G.; RICK, S.; COHEN, J. D. Neuroeconomics. *Annu. Rev. Psychol.*, v. 59, p. 647 – 672, 2008.
- [41] RANGEL, A.; CAMERER, C.; MONTAGUE, P. R. A framework for studying the neurobiology of value-based decision making. *Nat. Rev. Neurosci.*, v. 9, n. 7, p. 545 – 556, 2008.
- [42] FORREST, D.; SIMMONS, R.; CHESTERS, N. Buying a dream: Alternative models of demand for lotto. *Econ. Inq.*, v. 40, n. 3, p. 485 – 496, 2002.
- [43] SORNETTE, D. et al. Endogenous versus exogenous shocks in complex networks: An empirical test using book sale rankings. *Phys. Rev. Lett.*, v. 93, p. 228701, 2004.
- [44] WANG, K.-M.; THI, T.-B. N. Testing for contagion under asymmetric dynamics: Evidence from the stock markets between US and Taiwan. *Physica A*, v. 376, n. 0, p. 422 – 432, 2007.
- [45] BONESRØNNING, H. Peer group effects in education production: Is it about congestion? *J. Socio-Econ.*, v. 37, n. 1, p. 328 – 342, 2008.

- [46] CRANE, R.; SORNETTE, D. Robust dynamic classes revealed by measuring the response function of a social system. *Proc. Natl. Acad. Sci. U.S.A.*, v. 105, n. 41, p. 15649 – 15653, 2008.
- [47] CHI, L.-C. Contagion and competitive effects of plan confirmation of reorganization filings: Evidence from the Taiwan Stock Market. *Econ. Model.*, v. 26, n. 2, p. 364 – 369, 2009.
- [48] ONNELA, J.-P.; REED-TSOCHAS, F. Spontaneous emergence of social influence in online systems. *Proc. Natl. Acad. Sci. U.S.A.*, p. 18375 – 18380, 2010.
- [49] BREUER, R.; BRETTEL, M. Short- and long-term effects of online advertising: Differences between new and existing customers. *J. Interact. Mark.*, v. 26, n. 3, p. 155 – 166, 2012.
- [50] GILDEN, D.; THORNTON, T.; MALLON, M. $1/f$ noise in human cognition. *Science*, v. 267, n. 5205, p. 1837 – 1839, 1995.
- [51] HAUSDORFF, J. M.; PENG, C.-K. Multiscaled randomness: A possible source of $1/f$ noise in biology. *Phys. Rev. E*, v. 54, p. 2154 – 2157, 1996.
- [52] AKS, D. J.; ZELINSKY, G. J.; SPROTT, J. C. Memory across eye-movements: $1/f$ dynamic in visual search. *Nonlinear Dynamics Psychol. Life Sci.*, v. 6, n. 1, p. 1 – 25, 2002.
- [53] ORDEN, G. C. V.; HOLDEN, J. G.; TURVEY, M. T. Human cognition and $1/f$ scaling. *J. Exp. Psychol. Gen.*, v. 134, n. 1, p. 117 – 123, 2005.
- [54] KREISS, J. K. et al. Aids virus infection in Nairobi prostitutes. *New Engl. J. Med.*, v. 314, n. 7, p. 414 – 418, 1986.
- [55] MORRIS, M.; KRETZSCHMAR, M. Concurrent partnerships and the spread of HIV. *AIDS*, v. 11, n. 5, p. 641 – 648, 1997.
- [56] NEWMAN, M. E. J. Spread of epidemic disease on networks. *Phys. Rev. E*, v. 66, p. 016128, 2002.
- [57] RILEY, S. et al. Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions. *Science*, v. 300, n. 5627, p. 1961 – 1966, 2003.

- [58] LIPSITCH, M. et al. Transmission dynamics and control of severe acute respiratory syndrome. *Science*, v. 300, n. 5627, p. 1966 – 1970, 2003.
- [59] NEUMANN, G.; NODA, T.; KAWAOKA, Y. Emergence and pandemic potential of swine-origin H1N1 influenza virus. *Nature*, v. 459, n. 7249, p. 931 – 939, 2009.
- [60] CURRAN, J. W. et al. Epidemiology of HIV infection and AIDS in the United States. *Science*, 239, p. 610 – 616, 1988.
- [61] BOR, R.; MILLER, R.; GOLDMAN, E. HIV/AIDS and the family: a review of research in the first decade. *J. Fam. Ther.*, 15, n. 2, p. 187 – 204, 1993.
- [62] VOGL, D. et al. Symptom prevalence, characteristics, and distress in AIDS outpatients. *J. Pain Symptom Manage.*, 18, n. 4, p. 253 – 262, 1999.
- [63] HAYES, R.; WEISS, H. Understanding HIV epidemic trends in africa. *Science*, 311, n. 5761, p. 620 – 621, 2006.
- [64] FONSECA, M. G. P.; BASTOS, F. I. Twenty-five years of the AIDS epidemic in Brazil: principal epidemiological findings, 1980-2005. *Rev. Saude Publ.*, 23, p. S333 – S343, 2007.
- [65] WALKER, B. D.; BURTON, D. R. Toward an AIDS vaccine. *Science*, 320, n. 5877, p. 760 – 764, 2008.
- [66] HALL, H. I. et al. Epidemiology of HIV infection in large urban areas in the United States. *PLoS ONE*, 5, n. 9, p. e12756, 2010.
- [67] LEVY, J. Pathogenesis of human-immunodeficiency-virus infection. *Microbiol. Rev.*, v. 57, n. 1, p. 183 – 289, 1993.
- [68] AMUNDSEN, E. J. A. et al. Definition and estimation of an actual reproduction number describing past infectious disease transmission: application to HIV epidemics among homosexual men in Denmark, Norway and Sweden. *Epidemiol. Infect.*, 132, n. , p. 1139 – 1149, 2004.
- [69] UNAIDS. Disponível em: <http://www.unaids.org/globalreport/documents/20101123_GlobalReport_full_en.pdf>. Acesso em: setembro de 2012.

- [70] QUINN, T. C. Global burden of the HIV pandemic. *Lancet*, v. 348, n. 9020, p. 99 – 106, 1996.
- [71] MEINERS, C. et al. Modeling HIV/AIDS drug price determinants in Brazil: Is generic competition a myth? *PLoS ONE*, 6, n. 8, p. e23478, 2011.
- [72] WHO. Disponível em: <http://www.who.int/hiv/data/2011_epi_core_en.png>. Acesso em: setembro de 2012.
- [73] BAILEY, N. T. J. Simplified modelling of the population dynamics of HIV/AIDS. *J. R. Statist. Soc. A*, 151, n. , p. 31 – 43, 1988.
- [74] TEIXEIRA, P. R.; ANTÔNIO, M. V.; BARCAROLO, J. Antiretroviral treatment in resource-poor settings: the Brazilian experience. *AIDS*, 18, n. 3, p. 5 – 7, 2004.
- [75] DATASUS. Disponível em: <<http://www2.aids.gov.br/cgi/deftohtm.exe?tabnet/br.def>>. Acesso em: setembro de 2012.
- [76] NI, S. et al. Epidemic outbreaks in growing scale-free networks with local structure. *Physica A*, v. 387, n. 21, p. 5295 – 5302, 2008.
- [77] NI, S.; WENG, W.; ZHANG, H. Modeling the effects of social impact on epidemic spreading in complex networks. *Physica A*, v. 390, n. 23 – 24, p. 4528 – 4534, 2011.
- [78] GRANGEIRO, A.; ESCUDER, M. M. L.; CASTILHO, E. A. Magnitude and trend of the aids epidemic in brazilian cities, from 2002 to 2006. *Rev. Saude Publ.*, v. 44, n. 3, p. 430 – 441, 2010.
- [79] NISHIURA, H. Correcting the actual reproduction number: A simple method to estimate r_0 from early epidemic growth data. *Int. J. Environ. Res. Public Health*, 7, n. 1, p. 291 – 302, 2010.
- [80] GRAN, J. M. et al. Growth rates in epidemic models: Application to a model for HIV/AIDS progression. *Stat. Med.*, 27, n. 23, p. 4817 – 4834, 2008.
- [81] IBGE. Disponível em: <ftp://ftp.ibge.gov.br/Censos/Censo_Demografico_2000/Dados_do_Universo/Municipio>. Acesso em: setembro de 2012.

- [82] BETTENCOURT, L. M. A. et al. Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl. Acad. Sci. USA*, v. 104, n. 17, p. 7301–7306, 2007.
- [83] WEST, G. B.; BROWN, J. H.; ENQUIST, B. J. The fourth dimension of life: Fractal geometry and allometric scaling of organisms. *Science*, v. 284, n. 5420, p. 1677–1679, 1999.
- [84] NIKLAS, K. J. Plant allometry: is there a grand unifying theory? *Biol. Rev. Camb. Philos. Soc.*, v. 79, n. 4, p. 871 – 889, 2004.
- [85] NUNN, A. S. et al. Evolution of antiretroviral drug costs in Brazil in the context of free and universal access to AIDS treatment. *PLoS Med.*, v. 4, n. 11, p. e305, 2007.
- [86] PORCO, T. C. et al. Decline in HIV infectivity following the introduction of highly active antiretroviral therapy. *AIDS*, v. 18, n. 1, p. 81 – 88, 2004.
- [87] GRANGEIRO, A. et al. Late entry into HIV care: Estimated impact on AIDS mortality rates in Brazil, 2003 – 2006. *PLoS ONE*, 6, n. 1, p. e14585, 2011.
- [88] GREENHALGH, D.; DOYLE, M.; LEWIS, F. A mathematical treatment of AIDS and condom use. *IMA J. Math. Appl. Med. Biol.*, v. 18, n. 3, p. 225 – 262, 2001.
- [89] DUSHOFF, J. et al. Dynamical resonance can account for seasonality of influenza epidemics. *Proc. Natl. Acad. Sci. USA*, v. 101, n. 48, p. 16915 – 16916, 2004.
- [90] CHOWELL, G.; NISHIURA, H.; BETTENCOURT, L. M. A. Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. *J. R. Soc. Interface*, v. 4, n. 12, p. 155 – 166, 2007.
- [91] JIANG, D. et al. The long time behavior of DI SIR epidemic model with stochastic perturbation. *J. Math. Anal. Appl.*, v. 372, n. 1, p. 162 – 180, 2010.
- [92] ROCHA, L. E. C.; LILJEROS, F.; HOLME, P. Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts. *PLoS Comput. Biol.*, v. 7, n. 3, p. e1001109, 2011.
- [93] JUNIOR, S. d. P. et al. Spreading patterns of the influenza A (H1N1) pandemic. *PLoS ONE*, v. 6, n. 3, p. e17823, 2011.

- [94] KEELING, M. J.; EAMES, K. T. D. Networks and epidemic models. *J. R. Soc. Interface*, v. 2, p. 295 – 307, 2005.
- [95] LEVINS, R. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the ESA*, v. 15, n. 3, p. 237 – 240, 1969.
- [96] COHEN, J. More woes for struggling HIV vaccine field. *Science*, v. 340, n. 6133, p. 667, 2013.
- [97] MCKANE, A. J.; NEWMAN, T. J. Stochastic models in population biology and their deterministic analogs. *Phys. Rev. E*, v. 70, p. 041902, 2004.
- [98] ANTONOVICS, J.; MCKANE, A. J.; NEWMAN, T. J. Spatiotemporal dynamics in marginal populations. *Am. Nat.*, v. 167, n. 1, p. 16 – 27, 2006.
- [99] LI, C.; BOVENKAMP, R. van de; MIEGHEM, P. V. Susceptible-infected-susceptible model: A comparison of n -intertwined and heterogeneous mean-field approximations. *Phys. Rev. E*, v. 86, p. 026116, 2012.
- [100] CATOR, E.; MIEGHEM, P. V. Susceptible-infected-susceptible epidemics on the complete graph and the star graph: Exact analysis. *Phys. Rev. E*, v. 87, p. 012811, 2013.
- [101] KAMPEN, N. G. V. *Stochastic Processes in Physics and Chemistry*. 3rd. ed. [S.l.]: North Holland, 2007.
- [102] DAUXOIS, T. et al. Enhanced stochastic oscillations in autocatalytic reactions. *Phys. Rev. E*, v. 79, p. 036112, 2009.
- [103] MCKANE, A. J.; NEWMAN, T. J. Predator-prey cycles from resonant amplification of demographic stochasticity. *Phys. Rev. Lett.*, v. 94, p. 218102, 2005.
- [104] ANNA, P. de et al. Spatial model of autocatalytic reactions. *Phys. Rev. E*, v. 81, n. 5 Pt 2, p. 056110, 2010.
- [105] CIANCI, C. et al. Analytical study of non gaussian fluctuations in a stochastic scheme of autocatalytic reactions. *Eur. Phys. J.-Spec. Top.*, v. 212, n. 1, p. 5 – 22, 2012.

- [106] ROSS, T. M. Dengue virus. *Clin. Lab. Med.*, v. 30, n. 1, p. 149 – 160, 2010.
- [107] WHO | Dengue and severe dengue. Disponível em: <<http://www.who.int/mediacentre/factsheets/fs117/en/>>. Acesso em: agosto de 2013.
- [108] WHO | Better environmental management for control of dengue. Disponível em: <<http://www.who.int/heli/risks/vectors/denguecontrol/en/>>. Acesso em: agosto de 2013.
- [109] DENGUEVIRUSNET. Disponível em: <<http://www.denguevirusnet.com/life-cycle-of-aedes-aegypti.html>>. Acesso em: agosto de 2013.
- [110] GUZMAN, A.; ISTÚRIZ, R. E. Update on the global spread of dengue. *Int. J. Antimicrob. Ag.*, v. 36S, p. S40, 2010.
- [111] GHARBI, M. et al. Time series analysis of dengue incidence in Guadeloupe, French West Indies: Forecasting models using climate variables as predictors. *BMC Infect. Dis.*, v. 11, p. 166, 2011.
- [112] HU, W. et al. Spatial analysis of notified dengue fever infections. *Epidemiol. Infect.*, v. 139, p. 391 – 399, 2011.
- [113] WHO | Dengue. Disponível em: <<http://www.who.int/denguecontrol/en/>>. Acesso em: agosto de 2013.
- [114] DATASUS. Disponível em: <<http://dtr2004.saude.gov.br/sinanweb/index.php>>. Acesso em: agosto de 2013.
- [115] CHOWELL, G. et al. Spatial and temporal dynamics of dengue fever in Peru: 1994–2006. *Epidemiol. Infect.*, v. 136, p. 1667 – 1677, 2008.
- [116] BOTARI, T.; ALVES, S. G.; LEONEL, E. D. Explaining the high number of infected people by dengue in Rio de Janeiro in 2008 using a susceptible-infective-recovered model. *Phys. Rev. E*, v. 83, p. 037101, 2011.
- [117] STONE, L.; OLINKY, R.; HUPPERT, A. Seasonal dynamics of recurrent epidemics. *Nature*, v. 446, p. 533 – 536, 2007.

- [118] ZHANG, Q. et al. Comparison of detrending methods for fluctuation analysis in hydrology. *J. Hydrol.*, v. 400, n. 1–2, p. 121 – 132, 2011.
- [119] VECCHIO, A.; PRIMAVERA, L.; CARBONE, V. Periodic and aperiodic traveling pulses in population dynamics: An example from the occurrence of epidemic infections. *Phys. Rev. E*, v. 73, p. 031913, 2006.
- [120] ACUNA-SOTO, R.; VIBOUD, C.; CHOWELL, G. Influenza and pneumonia mortality in 66 large cities in the United States in years surrounding the 1918 pandemic. *PLoS ONE*, v. 6, n. 8, p. e23467, 2011.
- [121] MANTEGNA, R. N.; STANLEY, H. E. *Introduction to Econophysics: Correlations and Complexity in Finance*. 1. ed. [S.l.]: Cambridge University Press, 2000. ISBN 1139431226.
- [122] SPIEGEL, M. R. *Schaum's Mathematical Handbook of Formulas and Tables*. 2. ed. [S.l.]: McGraw-Hill, 1998. ISBN 0070382034.
- [123] Pearson Chi Square Test. Disponível em: <<http://reference.wolfram.com/mathematica/ref/PearsonChiSquareTest.html>>. Acesso em: junho de 2013.