

Beatriz de Rezende Barcellos Borges

Análise estocástica da propagação de uma doença de caráter epidemiológico.

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós–graduação em Engenharia Mecânica, do Departamento de Engenharia Mecânica da PUC-Rio.

Orientador : Profa. Roberta de Queiroz Lima Coorientador: Prof. Rubens Sampaio Filho

Rio de Janeiro Outubro de 2021



Beatriz de Rezende Barcellos Borges

Análise estocástica da propagação de uma doença de caráter epidemiológico.

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós–graduação em Engenharia Mecânica da PUC-Rio. Aprovada pela Comissão Examinadora:

Profa. Roberta de Queiroz Lima Orientador Departamento de Engenharia Mecânica – PUC-Rio

Prof. Rubens Sampaio Filho Coorientador Departamento de Engenharia Mecânica – PUC-Rio

Prof. Marcelo Amorim Savi Universidade Federal do Rio de Janeiro – UFRJ

Prof. Anderson Pereira Departamento de Engenharia Mecânica – PUC-Rio

Rio de Janeiro, 08 de Outubro de 2021

Todos os direitos reservados. A reprodução, total ou parcial do trabalho, é proibida sem a autorização da universidade, do autor e do orientador.

Beatriz de Rezende Barcellos Borges

Graduada em Engenharia Mecânica pela PUC-Rio. Trabalhou como pesquisadora, enquanto aluna, nas universidades UFRJ, no Rio de Janeiro, e Texas A & M University, Texas.

Ficha Catalográfica

Borges, Beatriz

Análise estocástica da propagação de uma doença de caráter epidemiológico. / Beatriz de Rezende Barcellos Borges; orientador: Roberta de Queiroz Lima; coorientador: Rubens Sampaio Filho. – 2021.

112 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Mecânica, 2021.

Inclui bibliografia

 Engenharia Mecânica – Teses. 2. Modelagem Estocástica. 3. Propagação de Incertezas. 4. Doenças Contagiosas.
 Vacinação. 6. Big Data. I. Lima, Roberta. II. Sampaio, Rubens. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Mecânica. IV. Título.

CDD: 620.11

Agradecimentos

Começo meus agradecimentos com uma menção especial aos meus orientadores Dra. Roberta Lima e Dr. Rubens Sampaio pelo conhecimento e entusiasmo sobre o assunto. Em particular, gostaria de agradecer à Roberta, que além de me ensinar sobre pesquisa e me orientar ao longo dessa jornada, foi de extrema sensibilidade e carinho nos momentos mais difíceis.

Agradeço à minha família, em especial à minha mãe, Elza, e ao meu pai, Victor, por me darem, desde sempre, o suporte para que eu pudesse caminhar sozinha e vencer minhas próprias batalhas. Obrigada pelo abrigo, amizade, conversas e conselhos nesses tempos difíceis que todos nós estamos enfrentando. Agradeço também à Bruna, pelas longas conversas que me trouxeram até aqui com a certeza dos passos que tenho dado.

Sou grata pelos meus colegas, amigos na verdade, de mestrado pelo carinho, acolhimento e ensinamentos passados ao longo desses dois anos. Obrigada por criarem, de perto ou de longe, um ambiente amigável e acolhedor para o desenvolvimento deste trabalho.

Gostaria de deixar aqui registrado meu agradecimento ao meu parceiro de vida, Pedro, por cada gesto de amor, acolhimento, paciência e incentivo. Um obrigada especial à família Savi, por me tratar como um dos seus e me acolher com carinho ao longo dessa jornada.

Gostaria de agradecer ao CNPQ e PUC-Rio pelo suporte financeiro ao longo do meu mestrado.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Resumo

Borges, Beatriz; Lima, Roberta; Sampaio, Rubens. Análise estocástica da propagação de uma doença de caráter epidemiológico.. Rio de Janeiro, 2021. 112p. Dissertação de Mestrado – Departamento de Engenharia Mecânica, Pontifícia Universidade Católica do Rio de Janeiro.

Este trabalho analisa a propagação de uma doença epidemiológica com uma abordagem estocástica. Na análise, o número de indivíduos que cada membro infectado da população pode infectar é modelado como uma variável aleatória e o número de indivíduos infectados ao longo do tempo é modelado como um processo estocástico de ramificação. O foco do trabalho é caracterizar a influência do modelo probabilístico da variável aleatória que modela o contágio entre indivíduos na disseminação da doença e na probabilidade de extinção, e analisar a influência de uma vacinação em massa no controle da propagação da doença. A comparação é feita com base em histogramas e estatísticas amostrais do número de indivíduos infectados ao longo do tempo, como média e variância. Os modelos estatísticos referentes à parte que trata de uma população não vacinada são calculados usando simulações de Monte Carlo para 3 diferentes famílias de variáveis aleatórias: binomial, geométrica-1 e geométrica-0. Para cada família, 21 distribuições diferentes foram selecionadas e, para cada distribuição, 4000 simulações do processo de ramificação foram computadas. Os modelos estatísticos referentes a uma população parcialmente vacinada foram calculados usando simulações de Monte Carlo para a família de variável aleatória binomial. Para essa família, 21 distribuições diferentes foram selecionadas e, para cada uma delas foram escolhidas 6 diferentes percentagens de população vacinada. Para cada percentagem, foram analisadas vacinas com 4 diferentes eficácias. No total, foram realizadas 2.2 milhões de simulações, caracterizando o problema como big data.

Palavras-chave

Modelagem Estocástica; Propagação de Incertezas; Doenças Contagiosas; Vacinação; Big Data.

Abstract

Borges, Beatriz; Lima, Roberta (Advisor); Sampaio, Rubens (Co-Advisor). Stochastic analyses of the spread of an epidemiological disease. Rio de Janeiro, 2021. 112p. Dissertação de Mestrado – Departamento de Engenharia Mecânica, Pontifícia Universidade Católica do Rio de Janeiro.

This work analyzes the spread of an epidemiological disease with a stochastic approach. In the analysis, the number of individuals that each infected member of the population can infect is modeled as a random variable and the number of infected individuals over time is modeled as a stochastic branching process. The focus of the work is to characterize the influence of the probabilistic model of the random variable that models contagion between individuals on the spread of the disease and the probability of extinction, and to analyze the influence of mass vaccination in controlling the spread of a disease. The comparison is based on histograms and sample statistics of the number of infected individuals over time, such as mean and variance. Statistical models for the chapter dealing with a vaccine free population are calculated using Monte Carlo simulations for 3 different families of random variables: binomial, geometric-1 and geometric-0. For each of the 3 families, 21 different distributions were selected and, for each distribution, 4000 simulations of the branching process were computed. Statistical models for a partially vaccinated population were calculated using Monte Carlo simulations for one family of random variable: the binomial. For it, 21 different distributions were selected and, for each of them, 6 different percentages of the vaccinated population were chosen. For each of them, 4 different vaccine efficacy were stipulated. In total, 2.2 million simulations were performed, featuring a big data problem.

Keywords

Stochastic Modeling; Uncertainties Propagation; Infectious Disease; Vaccination; Big Data.

Sumário

1 Introdução	14
1.1 O Que é Epidemiologia?	14
1.2 Classificação de Doenças Infecciosas	14
1.3 Modelagens da Propagação de uma Doença Contagiosa	15
1.4 Objetivos da Dissertação	16
1.5 Organização da Dissertação	17
2 Modelos Matemáticos Determinísticos Aplicados à Epide-	
miologia	19
2.1 Contexto Histórico	19^{-1}
2.2 Modelos de Compartimentos	20
2.2.1 Modelo SI	20
2.2.2 Modelo SIR	22
2.3 Número Básico de Reprodução	24
3 Conceitos Básicos de Probabilidade	27
3.1 Conceitos Básicos	27
3.1.1 Espaço de Probabilidade	27
3.1.2 Condicionamento	28
3.1.3 Independência	28
3.2 Variáveis Aleatórias Discretas	28
3.2.1 Definição	28
3.2.2 Algumas Estatísticas de Variáveis Aleatórias Discretas	29
3.3 Vetor de Variáveis Aleatórias Discretas	30
3.3.1 Definição	30
3.3.2 Algumas Estatísticas de Vetores Aleatórios	31
3.4 Soma de Variáveis Aleatórias	31
3.5 Processos Estocásticos	32
3.6 Método de Monte Carlo	32
3.6.1 Lei dos Grandes Números	33
3.6.2 Teorema do Limite Central	34
3.7 Quantificação e Propagação de Incertezas	34
4 Função Geradora de Probabilidade	35
4.1 Função Geradora	35
4.2 Função Geradora de Probabilidade - PGF	36
4.3 Cálculo de Média e Variância de Variáveis Aleatórias Através da	
Função Geradora de Probabilidade	37
4.4 Soma de Varáveis Aleatórias Independentes	38
5 Construção do Modelo Probabilístico para o Contágio e a	
Propagação de uma Doença	40
5.1 Construção do Modelo Probabilístico para o Contágio e a Propa-	
gação de uma Doença	40

5.2	Função Geradora de Probabilidade da Propagação de uma Doença	41
5.3	Função de Distribuição de Probabilidade do Total de Infectados	42
5.4	Algumas Estatísticas da Propagação	43
5.4.1	Média e Variância do Total de Infectados	43
5.5	Probabilidade de Extinção	45
6	Influência do Modelo do Contágio no Comportamento da	
	Doença	47
6.1	Estratégia para Analisar a Influência do Modelo Probabilístico do	
	Contágio na Propagação da Doença	47
6.2	Modelo Probabilístico de ${f B}$ - Binomial $({f m},{f p})$	48
6.2.1	Contágio $\mathbf{B} = $ binomial (3, 0.16), com $\mu_{\mathbf{B}} < 1$	50
6.2.2	2 Contágio $\mathbf{B} = \text{binomial} (3, 0.33), \text{ com } \mu_{\mathbf{B}} = 1$	51
6.2.3	B Contágio $\mathbf{B} = \text{binomial} (3, 0.50), \text{ com } \mu_{\mathbf{B}} > 1$	52
6.3	Modelo Probabilístico de ${f B}$ - Geométrica Parametrizada $({f p})$	53
6.3.1	Contágio $\mathbf{B} = \text{geométrica - 0 } (0.4), \text{ com } \mu_{\mathbf{B}} < 1$	54
6.3.2	2 Contágio $\mathbf{B} = \text{geométrica - 0 } (0.5), \text{ com } \mu_{\mathbf{B}} = 1$	55
6.3.3	B Contágio $\mathbf{B} = \text{geométrica - 0 } (0.6), \text{ com } \mu_{\mathbf{B}} > 1$	56
6.4	Modelo Probabilístico de ${\bf B}$ - Geométrica $({\bf p})$	58
7	Influência da Vacinação na Modelagem do Processo de	
	Ramificação	60
7.1	O Que São as Vacinas ?	60
7.2	Modelos Utilizados na Modelagem de uma Vacinação em Massa	61
7.3	Construção de um Modelo Estocástico para uma População Parci-	
	almente Vacinada	61
8	Influência da Vacinação no Comportamento da Epidemia	64
8.1	Influência da Vacinação	64
8.2	Influência da Porcentagem da População Vacinada	66
8.3	Influência da Eficácia da Vacina	68
8.3.1	Vacinando 50 % da População	69
8.3.2	2 Vacinando 75 % da População	72
9	Conclusão	76
Ref	erências bibliográficas	79
Δ	Artigos Publicados	83

Lista de figuras

Figura 2.1	Diagrama de bloco do modelo SI.	20
Figura 2.2	Comportamento dos grupos de susceptíveis e infectados	
ao long	go do tempo, para valores de $\beta = 3.34$ e $i_0 = 0.03$.	22
Figura 2.3	Diagrama de bloco do modelo SIR.	22
Figura 2.4	Comportamento dos grupos de susceptíveis, infectados perados ao londo do tempo, para valores de $\beta = 3.34$	
$\alpha = 1$	$\frac{1}{10} = 0.03$	24
$\alpha = 1.0$ Figure 2.5	Influência de r_{-} na percentagem de infectados ao longo	24
do tem	nnuencia de T_0 na porcentagem de nnectados ao longo $po.$	25
Figura 3.1	Esquematização de como obter estatísticas amostrais da	
respost	ta de um sistema que depende de parâmetros estocásticos,	
modela	ados como variáveis aleatórias discretas.	33
Figura 5.1	Uma realização do processo de ramificação que modela	
o núme	ema realização do processo de ramineação que modera pro do infoctados ao longo do goraç cos Considera so $i_0 = 1$	/11
0 nume	ero de infectados ao longo de gerações. Considera-se $i_0 = 1$	41
Figura 6.1	Estratégia para analisar a influência do modelo probabi-	
lístico	do contágio na propagação da doença.	48
Figura 6.2	Histogramas normalizados com o total de infectados nas	
geraçõe	es 2, 4, 6 e 21, construídos com 4000 realizações.	49
Figura 6.3	Comparação entre os histogramas normalizados cons-	
truídos	s com 4000 realizações e as funções de massa calculadas	
analiti	camente.	50
Figura 6.4	Contágio $B =$ binomial (3, 0.16), com $\mu_B < 1$.	50
6.4(a)	Estatísticas Amostrais.	50
6.4(b)	Probabilidade de Extinção.	50
Figura 6.5	Comparação entre os histogramas normalizados cons-	
truídos	s com 4000 realizações e as funções de massa calculadas	
analiti	camente.	51
Figura 6.6	Contágio B = binomial (3, 0.33), com μ_B = 1.	52
6.6(a)	Estatísticas Amostrais.	52
6.6(b)	Probabilidade de Extinção.	52
Figura 6.7	Comparação entre os histogramas normalizados cons-	
truídos	s com 4000 realizações e as funções de massa calculadas	
analitie	camente.	52
Figura 6.8	Contágio $B =$ binomial (3, 0.50), com $\mu_B > 1$.	53
6.8(a)	Estatísticas Amostrais.	53
6.8(b)	Probabilidade de Extinção.	53
Figura 6.9	Histogramas normalizados com o total de infectados nas	~ .
geraçõe	es 2, 4, 6 e 12, construidos com 4000 realizações.	54
Figura 6.10	Comparação entre os histogramas normalizados da geo-	
métrica	a - U (0.4) construidos com 4000 realizações e as funções	
de mas	ssa calculadas analiticamente.	55
Figura 6.11	Contágio $B = \text{geométrica} - 0 \ (0.4), \text{ com } \mu_B < 1.$	55

6.11(a)Estatísticas Amostrais.	55
6.11(bProbabilidade de Extinção.	55
Figura 6.12 Comparação entre os histogramas normalizados da geo-	
métrica - 0 (0.5) construídos com 4000 realizações e as funções	
de massa calculadas analiticamente.	56
Figura 6.13 Contágio $B =$ geométrica - 0 (0.5), com $\mu_B = 1$.	56
6.13(a)Estatísticas Amostrais.	56
6.13(bProbabilidade de Extinção.	56
Figura 6.14 Comparação entre os histogramas normalizados da geo-	
métrica - 0 (0.6) construídos com 4000 realizações e as funções	
de massa calculadas analiticamente.	57
Figura 6.15 Contágio $B =$ geométrica - 0 (0.6), com $\mu_B > 1$.	57
6.15(a Estatísticas Amostrais.	57
6.15(bProbabilidade de Extinção.	57
Figura 6.16 Histogramas normalizados com o total de infectados nas	
gerações 5, 10, 15 e 20, construídos com 4000 realizações.	58
Figura 6.17 Contágio $B = \text{geométrica} - 1 (0.88), \text{ com } \mu_B > 1.$	59
6.17(a Estatísticas Amostrais.	59
6.17(bProbabilidade de Extinção.	59
Figura 6.18 Contágio $B = \text{geométrica} - 1 \ (0.78), \text{ com } \mu_B > 1.$	59
6.18(aEstatísticas Amostrais.	59
6.18(bProbabilidade de Extinção.	59
Figura 6.19 Contágio $B = \text{geométrica} - 1 (0.68), \text{ com } \mu_B > 1.$	59
6.19(a Estatísticas Amostrais.	59
6.19(bProbabilidade de Extinção.	59
Figura 7.1 Esquema de avaliação de um possível novo infectado.	62
Figura 8.1 Gráficos construídos com 4000 realizações do comporta-	
mento sem vacina.	65
8.1(a)Média amostral para os 3 valores críticos de μ_C .	65
8.1(b)Probabilidade de extinção para os 3 valores críticos de μ_C .	65
Figura 8.2 Gráficos construídos com 4000 realizações do comporta-	
mento de um local com 20% da população vacinada com uma	
vacina com 50.4% de eficácia.	65
8.2(a)Média amostral para os 3 valores críticos de μ_{C} .	65
8.2(b)Probabilidade de extinção para os 3 valores críticos de μ_C .	65
Figura 8.3 Histogramas normalizados com o total de infectados nas	
gerações 2, 4, 6 e 21, construídos com 4000 realizações.	66
Figura 8.4 Histogramas normalizados com o total de infectados nas	
gerações 2, 4, 6 e 10, construídos com 4000 realizações.	67
Figura 8.5 Gráficos construídos com 4000 realizações do comporta-	•••
mento de um local com 2% da população vacinada com uma	
vacina com 50.4% de eficácia.	67
$8.5(a)$ Média amostral para os 3 valores críticos de μ_{C}	67
$8.5(b)$ Probabilidade de extinção para os 3 valores críticos de μ_{C}	67
Figura 8.6 Histogramas normalizados com o total de infectados nas	
gerações 2, 4, 6 e 21, construídos com 4000 realizações.	68

Figura 8.7 Gráficos construídos com 4000 realizações do comporta-	
mento de um local com 50% da população vacinada com uma	
vacina com 50.4% de eficácia.	68
8.7(a)Média amostral para os 3 valores críticos de μ_C .	68
8.7(b)Probabilidade de extinção para os 3 valores críticos de μ_C .	68
Figura 8.8 Histogramas normalizados com o total de infectados nas	
gerações 2, 4, 6 e 21, construídos com 4000 realizações.	69
Figura 8.9 Gráficos construídos com 4000 realizações do comporta-	
mento de um local com 75% da população vacinada com uma	
vacina com 50.4% de eficácia.	69
8.9(a)Média amostral para os 3 valores críticos de μ_C .	69
8.9(b)Probabilidade de extinção para os 3 valores críticos de μ_C .	69
Figura 8.10 Histogramas normalizados com o total de infectados nas	
gerações 2, 4, 6 e 21, construídos com 4000 realizações.	70
Figura 8.11 Gráficos construídos com 4000 realizações do comporta-	
mento de um local com 50% da população vacinada com uma	
vacina com 50.4% de eficácia.	70
8.11(aMédia amostral para os 3 valores críticos de μ_C .	70
8.11(b) robabilidade de extinção para os 3 valores críticos de μ_C .	70
Figura 8.12 Histogramas normalizados com o total de infectados nas	
gerações 2, 4, 6 e 21, construídos com 4000 realizações.	71
Figura 8.13 Gráficos construídos com 4000 realizações do comporta-	
mento de um local com 50% da população vacinada com uma	
vacina com 62.1% de eficácia.	71
8.13(aMédia amostral para os 3 valores críticos de μ_C .	71
8.13(b) robabilidade de extinção para os 3 valores críticos de μ_C .	71
Figura 8.14 Histogramas normalizados com o total de infectados nas	
gerações 2, 4, 6 e 21, construídos com 4000 realizações.	72
Figura 8.15 Gráficos construídos com 4000 realizações do comporta-	
mento de um local com 50% da população vacinada com uma	
vacina com 70.4% de eficácia.	72
8.15(a)Média amostral para os 3 valores críticos de μ_C .	72
8.15(b) robabilidade de extinção para os 3 valores críticos de μ_C .	72
Figura 8.16 Histogramas normalizados com o total de infectados nas	
gerações 2, 4, 6 e 21, construídos com 4000 realizações.	73
Figura 8.17 Gráficos construídos com 4000 realizações do comporta-	
mento de um local com 75% da população vacinada com uma	
vacina com 50.4% de eficácia.	73
8.17(a)Média amostral para os 3 valores críticos de μ_C .	73
8.17(b) robabilidade de extinção para os 3 valores críticos de μ_C .	73
Figura 8.18 Histogramas normalizados com o total de infectados nas	
gerações 2, 4, 6 e 21, construídos com 4000 realizações.	74
Figura 8.19 Gráficos construídos com 4000 realizações do comporta-	
mento de um local com 75% da população vacinada com uma	
vacina com 62.1% de eficácia.	74
8.19(a) Média amostral para os 3 valores críticos de μ_C .	74
8.19(b) Probabilidade de extinção para os 3 valores críticos de μ_C .	74

Figura 8.20 Histogramas normalizados com o total de infectados nas	
gerações 2, 4, 6 e 21, construídos com 4000 realizações.	75
Figura 8.21 Gráficos construídos com 4000 realizações do comporta-	
mento de um local com 75% da população vacinada com uma	
vacina com 70.4% de eficácia.	75
8.21(a)Média amostral para os 3 valores críticos de μ_C .	75
8.21(b)Probabilidade de extinção para os 3 valores críticos de μ_C .	75

Lista de tabelas

Tabela 2.1	Tabela com os valores dos parâmetros da simulação do	
modelo	SI.	21
Tabela 2.2	Tabela com os valores dos parâmetros da simulação do	
modelo	SIR.	23
Tabela 2.3	Tabela com os valores de r_o [16, 31].	25

1 Introdução

1.1 O Que é Epidemiologia?

Epidemiologia é uma área de estudos que trata da compreensão do processo saúde-doença no âmbito de populações. A palavra *epidemiologia*, do grego, significa "*epi*" sobre, "*demos*" povo, "*logos*" estudo, ou seja, estudo sobre povos/populações. O primeiro registro que se tem desta palavra é datado de 1802, em *Epidemiologia Española* [8].

De acordo com Olsen [28], "Epidemiologia é o estudo da distribuição e determinantes de estados ou eventos relacionados à saúde em populações específicas, e a aplicação deste estudo para o controle de problemas de saúde". Distribuição é o conceito utilizado para analisar a frequência e o padrão dos eventos de saúde em uma população. Determinantes, as causas e outros fatores que influenciam a ocorrência de doenças e eventos relacionados à saúde.

Resumindo, epidemiologia é o estudo (científico, sistemático, baseado em dados) da distribuição (frequência, padrão) e determinantes (causas, fatores de risco) de estados e eventos relacionados à saúde em populações específicas, e a aplicação deste estudo para o controle de problemas de saúde [10].

Neste trabalho vamos focar no estudo da transmissão de doenças infecciosas, abordar a influência do comportamento da sociedade na propagação da doença e a aplicação de métodos de contenção de transmissão de doenças, e as consequências desses fatores na sociedade.

1.2 Classificação de Doenças Infecciosas

As doenças infecciosas podem ser causadas por diversos agentes patogênicos, como vírus, fungos, bactérias, parasitas ou toxinas, como descrito por Martcheva [25]. Além de distinguir as doenças infecciosas pela causa, é possível diferenciá-las pelo tipo de transmissão [7, 13, 25].

 Transmissão direta: infecção adquirida por contacto com um indivíduo doente.

- Transmissão indireta: infecção adquirida através da do consumo de água, alimentos ou por contato com outro agente infectante.
- Transmissão aérea: adquirida através do ar e dos agentes infectantes nele contidos.
- Transmissão vertical: transmitida através da placenta, de mãe para filho(a).
- Transmissão critogênica: infecção de porta de entrada desconhecida.
- Transmissão endógena: infecção devido a um micro-organismo já existente no organismo, e que, por qualquer razão, se torna patogênico.
- Transmissão exógena: provocada por micro-organismos provenientes do exterior, como alimentos e água contaminada.
- Transmissão por vetor: transmitida por um vetor, como mosquito, moluscos, etc.

É possível que algumas doenças tenham mais de uma via de transmissão. A influenza pode, por exemplo, ser transmitida por contaminação exógena, através de animais, e transmissão aérea, devido aos agentes infectantes no ambiente. Já HIV, por sua vez, pode ser transmitida verticalmente, como no nascimento, transmitida diretamente através do ato sexual, além de ser transmitida através de transfusão de sangue e compartilhamento de material infectado [26]. A transmissão de doenças é um dos principais pontos dos estudos epidemiológicos.

1.3 Modelagens da Propagação de uma Doença Contagiosa

As primeiras contribuições para a análise epidemiológica matemática do século XX foram baseadas em modelos compartimentados, propostos por Ross, Hamer, McKendrick e Kermack [22, 20]. Esses modelos, conhecidos como modelos de compartimentos, dividem a população de um determinado local em diferentes grupos. A quantidade de grupos é definida de acordo com a complexidade escolhida. Dois dos modelemos mais utilizados são o modelo SI e o modelo SIR, que serão detalhados nos capítulos a seguir.

Esses modelos propostos por McKendrick e Kermack, assim como alguns outros modelos de compartimentos modelam seus parâmetros de forma determinística. Isso implica que cada um dos indivíduos infectados com a doença infectam um mesmo número de indivíduos. No entanto, uma transmissão de uma doença se da de forma estocástica. Enquanto um indivíduo contaminado pode passar a doença para 10 outros indivíduos, um segundo indivíduo pode não contaminar ninguém. Algumas modelagens misturam as duas abordagens. Utilizam métodos estocásticos para definir a transmissão de uma doença entre os indivíduos, reduzem esta análise a uma estatística, e a utilizam como as taxas de transmissão ou recuperação para analisar o comportamento do total de infectados ao longo do tempo [9, 37, 38].

Este trabalho analisa essa transmissão sob a ótica estocástica, bem como analisa, sob a mesma abordagem, os efeitos que isso tem na propagação de uma doença contagiosa em meio a uma população.

1.4 Objetivos da Dissertação

Na primeira parte do trabalho é feita uma análise determinística do processo de propagação de uma doença. Para tal, são utilizados os modelos de compartimentos [22]. O primeiro modelo de compartimentos estudado será o modelo SI, composto apenas por dois compartimentos (susceptíveis e infectados). O segundo modelo de compartimentos será o SIR. Os parâmetros determinísticos destes modelos são β , taxa de contágio; α , taxa de recuperação; e r_o , número básico de reprodução, razão entre $\beta \in \alpha$.

Na segunda parte do trabalho o problema é abordado com uma modelagem estocástica. O número de indivíduos infectados por cada membro infectado da população é modelado por uma variável aleatória discreta e o número de indivíduos infectados ao longo do tempo é modelado por um processo estocástico de ramificação. Utilizando o método de Monte Carlo para caracterizar a propagação das incertezas na disseminação da doença, serão construídos histogramas do número de indivíduos infectados. Estatísticas amostrais e a probabilidade de extinção para as três famílias diferentes de variáveis aleatórias (v.a.) também serão calculadas. A influência do modelo probabilístico que modela o contágio entre os indivíduos no processo de ramificação será feita através da comparação destas estatísticas e histogramas. Vale ressaltar que a utilização do método de Monte Carlo caracteriza o problema abordado como um problema de *Big Data*.

Neste trabalho também será feito um estudo sobre as soluções analíticas deste problema. A partir das funções geradoras apresentadas obteremos as médias, variâncias e probabilidades de extinção analíticas. Bem como obteremos a função de massa do total de infectados em cada geração. Além disto, será feita a comparação entre os resultados amostrais e analíticos com o objetivo de validar a modelagem proposta.

A terceira e última parte traz a propagação de uma doença em uma população vacinada. Nela temos o contato entre indivíduos capaz de transmitir uma carga viral suficiente para um contágio modelado por uma v.a. discreta, a porcentagem da população vacinada modelada por outra v.a discreta, e a eficácia da vacina modelada por uma terceira v.a. discreta. O número total de indivíduos infectados ao longo do tempo é modelado por um processo estocástico de ramificação. Para caracterizar a propagação das incertezas na disseminação da doença, construiremos histogramas do número de indivíduos infectados. Estatísticas amostrais também serão calculadas. A influência da eficácia da vacina no processo de ramificação, bem como da influência da porcentagem da população vacinada serão abordados neste trabalho junto com o cálculo destas estatísticas e histogramas. A probabilidade de extinção da epidemia será um dos pontos abordados nesta parte.

1.5 Organização da Dissertação

Esta dissertação está organizada da seguinte maneira: este primeiro capítulo apresenta uma introdução sobre o tema, objetivos da dissertação e a organização de cada capítulo.

O segundo capítulo traz os modelos matemáticos determinísticos que são aplicados à epidemiologia. Ele apresenta os modelos de compartimentos que foram utilizados para a análise inicial deste trabalho.

No terceiro capítulo estão descritos alguns dos conceitos básicos de probabilidade, fundamentais para o entendimento deste trabalho.

O quarto capítulo apresenta o conceito de função geradora de probabilidade (PGF). Este conceito é um dos pilares das análises feitas ao longo desta dissertação.

O quinto capítulo traz a construção do modelo probabilístico pra o contágio e para a propagação de uma doença dentro de uma sociedade. Além de trazer as escolhas dos parâmetros modelados, este capítulo apresenta os cálculos analíticos para a construção de algumas das estatísticas da propagação de uma doença.

O sexto capítulo mostra alguns dos resultados obtidos, evidenciando a influência do modelo do contágio no comportamento da doença ao longo do tempo.

O sétimo capítulo traz a influência da vacinação na modelagem do processo de ramificação. Neste capítulo são apresentadas alguns modelos já utilizados para modelar a vacinação em massa em uma sociedade e o modelo estocástico proposto para modelar o comportamento de uma epidemia em uma população parcialmente vacinada. No oitavo capítulo é apresentada a segunda bateria de resultados obtidos. Ele apresenta a influência da vacinação sob duas frentes: como a vacinação com uma vacina de diferentes eficácias influenciam a propagação de uma doença; e como a porcentagem da população vacinada também interfere nessa propagação.

O nono e último capítulo apresenta as conclusões obtidas a partir dos resultados descritos ao longo desta dissertação.

2 Modelos Matemáticos Determinísticos Aplicados à Epidemiologia

Este capítulo trata de alguns dos modelos epidemiológicos determinísticos mais difundidos na literatura: os modelos de compartimentos.

2.1 Contexto Histórico

Um dos primeiros estudos matemáticos sobre doenças infecciosas foi o trabalho de John Graunt [18], em 1662, sobre possíveis causas dos padrões de mortalidade em Londres, onde morava. Graunt analisou os registros de mortalidade antes da Grande Peste e forneceu um método para estimar o risco de morte por diversas doenças. Ele relatou tendências temporais para diversas doenças e apresentou uma das primeiras tábuas de vida (tabela utilizada principalmente no cálculo atuarial, em planos de previdência e seguros de vida, para calcular as probabilidades de vida e morte de uma população, em função da idade).

Por volta do século XVIII, ao estudar sobre a varíola, Bernoulli desenvolveu o primeiro modelo matemático epidemiológico [2]. Sua abordagem foi direcionada a questão dos riscos concorrentes: quando observa-se a morte causada por um agente k é impossível observar a morte do mesmo indivíduo por qualquer outro fator de mortalidade. Esta publicação influenciou publicações mais complexas (vide [3, 12]).

No século XVIII, durante um surto de epidemia de cólera em Londres, John Snow conseguiu identificar a fonte de infecção como uma das bombas de água da cidade [35]. Na mesma época, Willian Farr [14] propôs a lei que fundamentou o crescimento e a queda de epidemias: "Ascensão rápida no inicio, elevação lenta até o ápice e, em seguida uma queda mais rápida".

As primeiras contribuições para a análise epidemiológica matemática do século XX foram baseadas em modelos compartimentados, propostos por Ross, Hamer, McKendrick e Kermack [22, 20].

2.2

Modelos de Compartimentos

Um dos primeiros modelos de matemáticos desenvolvidos para a análise de epidemias foi o proposto por Kermack e McKendric [22]. Esse modelo, conhecido como modelo de compartimentos, divide a população de um determinado local em diferentes grupos. A quantidade de grupos é definida de acordo com o a complexidade escolhida.

Dois dos principais modelos estão descritos a seguir.

2.2.1 Modelo SI

O modelo SI é um modelo de compartimentos composto apenas por dois compartimentos: o primeiro, s, representa a porcentagem da população susceptível a doença. O segundo compartimento representa a porcentagem de indivíduos que foram infectados pela doença, dado por i. A relação entre os grupos $s \in i$ é dado pelo contágio entre os indivíduos, representado por β , como mostra a figura 2.1.



Figura 2.1: Diagrama de bloco do modelo SI.

Esse modelo supõem que o total de indivíduos deste local não muda com o decorrer da propagação da doença. Dito isto, o vínculo entre esses dois grupos é dado pela eq. 2-1 a seguir.

$$s(t) + i(t) = 1,$$
 (2-1)

para qualquer instante t.

A dinâmica do sistema composto por esses dois compartimentos é dada pelo problema de valor inicial (PVI) da eq. 2-2:

$$\begin{cases} s'(t) = -\beta \ s(t)i(t) \\ i'(t) = \beta \ s(t)i(t) \end{cases} \begin{cases} s(0) = s_0 & | \ s_0 \in (0,1) \\ i(0) = i_0 & | \ i_0 \in (0,1), \end{cases}$$
(2-2)

Substituindo o vínculo dado na equação 2-1 na segunda equação de 2-2, reduz-se o número de graus de liberdade do sistema, transformando um problema de duas equações em um problema de valor inicial, com apenas uma equação, como descrito a eq. 2-3:

$$\begin{cases} i'(t) = \beta \ (1 - i(t))i(t) \\ i(0) = i_0. \end{cases}$$
(2-3)

Resolvendo o problema de valor inicial chegamos à eq. 2-4,

$$i(t) = \frac{i_0}{i_0 + (1 - i_0)e^{-\beta t}}, \qquad (2-4)$$

a qual rege o comportamento da porcentagem de indivíduos infectados ao longo do tempo.

A fim de caracterizar o comportamento no início do contágio, avaliamos a equação 2-4 no tempo, em $t \approx 0$. Considerando no denominador $i_0 \approx 0$ e $(1 - t_0) \approx 1$, temos

$$i(t) \approx i_0 e^{\beta t}.\tag{2-5}$$

Com isso, é possível notar que temos um crescimento exponencial dos infectados logo no inicio da epidemia, como proposto por Farr [14], regido pelo valor do contágio β .

Para analisar o comportamento a longo prazo, isto é, quando $t \to \infty$, foi feita uma simulação numérica no Matlab utilizando o método de integração Runge-Kutta de quarta e quinta ordem (função *ode*45). Os parâmetros do modelo, como s_0 , $i_0 \in \beta$, e os parâmetros de integração, passo e tempo, estão descritos na tabela 2.1.

Parâmetros	Valores
$\overline{s_0}$	0.97
i_0	0.03
β	3.34
passo de integração h	$0.05~{\rm s}$
tempo de integração t	$15.00~\mathrm{s}$

Tabela 2.1: Tabela com os valores dos parâmetros da simulação do modelo SI.

A figura 2.2 mostra a variação da porcentagem de infectados e susceptíveis a doença ao longo do tempo. É possível observar nesta mesma figura que, com esse modelo, a população tende a ser 100% infectada quando t tende a infinito.

Esse modelo é conhecido como modelo logístico e tem como característica marcante a tendência a infecção de todos os indivíduos da população quando o tempo t tende a infinito.



Figura 2.2: Comportamento dos grupos de susceptíveis e infectados ao longo do tempo, para valores de $\beta = 3.34$ e $i_0 = 0.03$.

2.2.2 Modelo SIR

O modelo SIR proposto por Kermack e McKendric [22] é composto por três compartimentos: o primeiro, s, representa a porcentagem da população susceptível a doença. O segundo compartimento representa a porcentagem de indivíduos que foram infectados pela doença, dado por i. O terceiro grupo é o referente à porcentagem da população que não pode mais ser infectada, seja porque se recuperou e adquiriu os anticorpos necessários ou porque veio a óbito. Esse grupo é denominado recuperados e são representados pela variável r. A relação entre os grupos $s \in i$ é dado pela taxa contágio entre os indivíduos, representado por β . A relação dada ente $i \in r$ é representada pela taxa de recuperação da doença, representada por α , como mostra a figura 2.3.



Figura 2.3: Diagrama de bloco do modelo SIR.

Esse modelo também supõem que o total de indivíduos deste local não muda com o decorrer da propagação da doença. Dito isto, o vínculo entre esses três grupos é dado pela eq. 2-6 a seguir:

$$s(t) + i(t) + r(t) = 1,$$
 (2-6)

válido para qualquer instante t.

A dinâmica entre esses grupo é dada pelo PVI da eq. 2-7:

Capítulo 2. Modelos Matemáticos Determinísticos Aplicados à Epidemiologia23

$$\begin{cases} s'(t) = -\beta \ s(t)i(t) \\ i'(t) = \beta \ s(t)i(t) - \alpha \ i(t) \\ r'(t) = \alpha \ i(t) \end{cases} \begin{cases} s(0) = s_0 & | \ s_0 \in (0, 1), \\ i(0) = i_0 & | \ i_0 \in (0, 1), \\ r(0) = r_0 & | \ r_0 \in (0, 1), \end{cases}$$
(2-7)

onde β descreve o contágio da doença e α a recuperação dos infectados.

Dividindo a equação de s'(t) pela de r'(t) da eq. 2-7 temos:

$$\frac{ds}{dr} = -\frac{\beta}{\alpha}s(t). \tag{2-8}$$

Pela equação 2-8 vemos que o número de indivíduos susceptíveis à doença decai com o tempo de acordo com a relação entre β e α . Essa relação é conhecida como número básico de reprodução, representado por r_o .

Para analisar o comportamento a longo prazo, isto é, quando $t \to \infty$, foi feita uma simulação numérica no Matlab utilizando o método de integração Runge-Kutta de quarta e quinta ordem (função *ode*45). Os parâmetros do modelo, s_0 , i_0 , r_0 , $\beta \in \alpha$, e os parâmetros de integração, passo e tempo, estão descritos na tabela 2.2.

Parâmetros	Valores	Parâmetros	Valores
$\overline{s_0}$	0.97	β	3.34
i_0	0.03	α	1.00
r_0	0.00	r_o	3.34
passo de integração \boldsymbol{h}	$0.05~{\rm s}$	tempo de integração t	$15.00~\mathrm{s}$

Tabela 2.2: Tabela com os valores dos parâmetros da simulação do modelo SIR.

A figura 2.4 mostra a variação da porcentagem de susceptíveis, infectados e recuperados da doença ao longo do tempo. É possível observar nesta mesma figura que o total de infectados tem uma "ascensão rápida no inicio, elevação lenta até o ápice e, em seguida uma queda mais rápida", como enunciado por Farr [14].

O pico de infectados é determinado por uma conjunção de fatores como a porcentagem inicial de infectados, a taxa de transmissão, a taxa de recuperação e, por consequência, pelo r_o . A variação e determinação desses parâmetros é a chave da modelagem de compartimentos.



Figura 2.4: Comportamento dos grupos de susceptíveis, infectados e recuperados ao londo do tempo, para valores de $\beta = 3.34$, $\alpha = 1.00$, $i_0 = 0.03$.

Este modelo é a base para modelos de compartimentos mais complexos como os utilizados em [22, 31, 33].

2.3 Número Básico de Reprodução

O número básico de reprodução é um dos parâmetros mais importantes dentro da epidemiologia. Serve como base para a adoção de algumas medidas de controle adotadas pelos governos [11].

O número básico de reprodução de uma infecção é o número esperado de casos gerados diretamente por um indivíduo infectado em uma população na qual todos os indivíduos são suscetíveis à infecção. A definição pressupõe que nenhum outro indivíduo esteja infectado ou imunizado (naturalmente ou por vacinação). Como dito na seção 2.2.2, r_o pode ser dado pela razão entre os parâmetros que modelam o contágio β e a recuperação α , como mostra a equação 2-9, β

$$r_o = \frac{\beta}{\alpha}.\tag{2-9}$$

O valor de r_0 de uma determinada doença pode ser estimado através de diferentes métodos, como por exemplo, através da taxa de crescimento inicial das epidemias, da análise de equilíbrio da pandemia, dos processos que constituem a contaminação ou através da calibragem de parâmetros de epidemias passadas [26].

Cada método resulta em uma estimativa diferente. Por isso, encontramse diversos valores para o mesmo parâmetro. Esses valores, em geral, ficam dentro de uma faixa ou intervalo. Alguns desses intervalos para r_o encontrados na literatura para diferentes doenças estão descritos na tabela 2.3.

Doença	r_o
Influenza	0.9 - 2.8
Ebola	1.5 - 1.9
Gripe Sazonal	2.0 - 3.0
HIV	2.0 - 5.0
COVID - 19	1.9 - 5.7
Varíola	3.5 - 6.0
Rubéola	5.0 - 7.0
Sarampo	12 - 18

Tabela 2.3: Tabela com os valores de r_o [16, 31].

Para mostrar a influência que o valor de r_o têm no comportamento das curvas de infectados, foram feitas algumas simulações numéricas. Com o auxílio do *Matlab*, utilizando a função integradora *ode45* na eq. 2-7, foram calculados diferentes cenários. Cada cenário foi definido com um valor diferente de r_o , e tem como demais parâmetros os valores da tabela 2.2.



Figura 2.5: Influência de r_o na porcentagem de infectados ao longo do tempo.

A figura 2.5 mostra a influência de r_o na porcentagem de infectados ao longo do tempo. Supondo que o valor de α seja o mesmo para os diferentes valores de r_o . Quanto maior é o r_o maior é a taxa de contágio β . Os diferentes contágios geram comportamentos distintos na propagação da doença.

Métodos distintos de determinação do r_o podem resultar em diferentes contágios e, consequentemente, diferentes propagações da doença no meio da sociedade. A partir disto, surge a oportunidade de olhar o contágio e a propagação sob uma ótica estocástica.

Capítulo 2. Modelos Matemáticos Determinísticos Aplicados à Epidemiologia26

Vale destacar que não é só o fato de os resultados dos diferentes métodos de determinação resultarem em diferentes valores de r_o que abrem margem para utilizarmos uma modelagem estocástica. O contágio em si também é randômico. Os indivíduos infectados com uma doença contagiosa podem ou não infectar o mesmo número de indivíduos. Enquanto um pode contaminar 5 novos indivíduos, um outro pode não contaminar ninguém.

Nos capítulos a seguir serão apresentados alguns conceitos básicos de probabilidade necessários para a construção de um modelo probabilístico para modelar o contágio e a propagação de uma doença.

3 Conceitos Básicos de Probabilidade

Nos capítulos anteriores foram apresentados diversos modelos matemáticos que descrevem o comportamento da disseminação de uma doença contagiosa. No entanto, os modelos citados não capturam a estocasticidade inerente a um processo de contaminação. Por isso, se faz presente a oportunidade de analisar este problema através de uma abordagem estocástica.

Este capítulo apresenta alguns conceitos de probabilidade utilizados nos modelos estocásticos apresentados nesta dissertação.

3.1 Conceitos Básicos

3.1.1 Espaço de Probabilidade

Um espaço de probabilidade é definido por $(\Omega, \mathcal{F}, \mathbb{P}_r)$, tal que

- espaço amostral (Ω): conjunto não vazio de todos os resultados possíveis de um experimento,
- $-\sigma$ -álgebra em $\Omega(\mathcal{F})$: um conjunto de todos os eventos que podem ocorrer como consequências deste experimento. Isto é,

$$- \oslash \in \mathcal{F}; - A \in \mathcal{F} \implies A^C \in \mathcal{F}; - A_1, A_2, ..., \in \mathcal{F} \Longrightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F};$$

– medida de probabilidade em \mathcal{F} (\mathbb{P}_r): uma avaliação da probabilidade desses eventos, dado entre [0, 1], tal que:

$$-\mathbb{P}_r(\Omega) = 1 \in \mathbb{P}_r(\emptyset) = 0;$$

- dados $A_1, A_2, ...,$ eventos disjuntos em \mathcal{F} , então:

$$\mathbb{P}_r(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}_r(A_i).$$
(3-1)

Vale dizer que só é possível associar uma probabilidade a um evento que pertence a \mathcal{F} .

3.1.2 Condicionamento

Cada evento é uma informação. A cada informação é possível definir um novo espaço de probabilidade.

Temos A e $C \in \mathcal{F}$, com probabilidades $\mathbb{P}_r(A) \neq 0$ e $\mathbb{P}_r(C) \neq 0$.

Dado A, definimos o primeiro espaço de probabilidade: $(\Omega, \mathcal{F}, \mathbb{P}_r)$. Com a nova informação C, definimos o novo espaço de probabilidade $(C, \mathcal{F}_C, \mathbb{P}_{r_C})$, tal que \mathbb{P}_{r_C} é uma medida de probabilidade em \mathcal{F} . Para definirmos a probabilidade de A no novo espaço C temos:

$$\mathbb{P}_{r_C}(A) = \frac{\mathbb{P}_r(A \cap C)}{\mathbb{P}_r(C)} = \mathbb{P}_r(A|C), \qquad (3-2)$$

tal que $\mathbb{P}_r(A|C)$ é a probabilidade do evento A ocorrer, dado que o evento C ocorreu.

3.1.3 Independência

Dados dois eventos $A \in C$. Esses eventos são independentes se a ocorrência de um deles não interfere na probabilidade do outro evento acontecer. Ou seja, dados $\mathbb{P}_r(A)$, $\mathbb{P}_r(C) > 0$,

$$\mathbb{P}_r(A|C) = \mathbb{P}_r(A) \quad \text{e} \quad \mathbb{P}_r(C|A) = \mathbb{P}_r(C). \tag{3-3}$$

Se temos

$$\mathbb{P}_r(A|C) = \frac{\mathbb{P}_r(A \cap C)}{\mathbb{P}_r(C)} = \mathbb{P}_{r_C}(A) = \mathbb{P}_r(A),$$
(3-4)

então temos que os eventos $A \in C$ são independentes, e a probabilidade de acontecerem ao mesmo tempo é dada por $\mathbb{P}_r(A \cap C) = \mathbb{P}_r(A)\mathbb{P}_r(C)$.

3.2 Variáveis Aleatórias Discretas

3.2.1 Definição

Uma variável aleatória (v.a.) descreve um experimento. Uma v.a. discreta real X em $(\Omega, \mathcal{F}, \mathbb{P}_r)$ é uma função real [32] dada por:

$$\begin{aligned} X : \Omega &\mapsto \mathbb{R} \\ \omega &\to X(\omega), \end{aligned}$$

tal que sua imagem $X(\Omega)$ é um subconjunto discreto de \mathbb{R} , e

$$X^{-1}(x) = \{ \omega \in \Omega : X(\omega) = x \} \in \mathcal{F}, \ \forall x \in \mathbb{R},$$

permite o transporte da medida de probabilidade \mathbb{P}_r para a reta (chamada de mensurabilidade).

A função de massa de uma v.a. X é definida como a função que permite atribuir uma certa probabilidade a cada distinto valor X de uma v.a. discreta. Isto é, $p_X(x) = \mathbb{P}_r(X = x)$ para qualquer valor x resultante de uma v.a. Xque tome valores num conjunto discreto. Assim,

- $\sum_{x \in ImX} p_X(x_i) = 1$, sendo x_i os valores que a variável X pode assumir; - 0 < $p_X(x)$ < 1;
- $-p_X(x_i) = 0$, se x_i não for resultado da v.a. X.

3.2.2 Algumas Estatísticas de Variáveis Aleatórias Discretas

A média, ou valor esperado, é um objeto determinístico, que determina a melhor aproximação de uma variável aleatória X por uma constante. O valor esperado de uma v.a. pode ser calculado através da equação 3-5.

$$\mu_X = \mathbb{E}[X] = \sum_{x \in \operatorname{Im}X} x \mathbb{P}_r(X = x), \qquad (3-5)$$

quando essa soma converge absolutamente, isto é, $\sum_{x} |x \mathbb{P}_r(X = x)| < \infty$.

Variância é a norma do erro quadrático da aproximação feita pela média, erro absoluto. Em outras palavras é uma medida de dispersão em torno da média. Dada por:

$$\sigma_X^2 = \mathbb{E}[(X - \mathbb{E}(X))^2] = \sum_{x \in \mathrm{Im}X} (x - \mu_X)^2 \,\mathbb{P}_r(X = x), \tag{3-6}$$

quando essa soma converge.

Algumas das v.a. discretas abordadas neste trabalho estão listadas abaixo, com os respectivos valores de média, variância amostral, suporte e função de massa.

- Sendo X uma v.a. Bernoulli (p):
 - Suporte: $\{0, 1\}.$

- Função de Massa: $p_X(x) = \mathbb{P}_r(X = x) = \begin{cases} p & \text{se } x = 1\\ 1-p & \text{se } x = 0. \end{cases}$

- Esperança: $\mathbb{E}[X] = p$.
- Variância: $\sigma_X^2 = p(1-p)$.
- Sendo X uma v.a. Binomial (m, p):

- Suporte: $\{0, 1, 2, \dots, m\}$.
- Função de Massa: $p_X(x) = \mathbb{P}_r(X = x) = \frac{m!}{x!(m-x)!} p^x (1-p)^{(m-x)}.$
- Esperança: $\mathbb{E}[X] = mp$.
- Variância: $\sigma_X^2 = mp(1-p)$.
- Sendo X uma v.a. Geométrica 0 (p):
 - Suporte: $\{0, 1, 2, 3, \dots\}$.
 - Função de Massa: $p_X(x) = \mathbb{P}_r(X = x) = (1-p)^{x-1}p.$

- Esperança:
$$\mathbb{E}[X] = \frac{1-p}{p}$$
.
- Variância: $\sigma_X^2 = \frac{2(p-1)^2 - p}{n^2}$.

- Sendo X uma v.a. Geométrica 1 (p):
 - Suporte: $\{1, 2, 3, \dots\}$.
 - Função de Massa: $p_X(x) = \mathbb{P}_r(X = x) = (1-p)^{x-1}p.$

- Esperança:
$$\mathbb{E}[X] = \frac{1}{p}$$

- Variância:
$$\sigma_X^2 = \frac{(1-p)}{p^2}$$
.

3.3 Vetor de Variáveis Aleatórias Discretas

3.3.1 Definição

Dados X e Y, variáveis aleatórias discretas em $(\Omega, \mathcal{F}, \mathbb{P}_r)$, a função de massa conjunta $p_{X,Y}$ de X e Y é a função $p_{X,Y} : \mathbb{R}^2 \to [0, 1]$, definida por

$$p_{X,Y}(x,y) = \mathbb{P}_r\Big(\{\omega \in \Omega : X(\omega) = x \in Y(\omega) = y\}\Big), \tag{3-7}$$

usualmente representada por $p_{X,Y}(x,y) = \mathbb{P}_r(X = x, Y = y).$

Com isso, temos que

$$\sum_{x \in \text{Im } X} \sum_{y \in \text{Im } Y} p_{X,Y}(x,y) = 1,$$
(3-8)

onde Im X é a imagem de X e Im Y é a imagem de Y.

As funções de massa marginais de X e Y podem ser achadas através da função de massa conjunta, como mostrado a seguir:

$$p_X(x) = \mathbb{P}_r(X = x) = \sum_y p_{X,Y}(x, y),$$
 (3-9)

$$p_Y(y) = \mathbb{P}_r(Y = y) = \sum_x p_{X,Y}(x, y).$$
 (3-10)

Estendendo esse conceito para um conjunto de n variáveis discretas é possível dizer que, dado um vetor $\mathbf{X} = (X_1, X_2, X_3, ..., X_n)$, sua função de massa conjunta é dada por:

$$p_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}_r(X_1 = x_1, X_2 = x_2, ..., X_n = x_n)$$
(3-11)

para $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$.

3.3.2 Algumas Estatísticas de Vetores Aleatórios

Dado as mesmas v.a. $X \in Y \in (\Omega, \mathcal{F}, \mathbb{P}_r)$, o valor esperado do vetor aleatório pode ser calculado diretamente a partir da função de massa conjunta $p_{X,Y}(x, y)$, tal que:

$$\mathbb{E}[XY] = \sum_{x \in \operatorname{Im} X} \sum_{y \in \operatorname{Im} Y} (x, y) \mathbb{P}_r \Big(X = x, \ Y = y \Big),$$
(3-12)

válido sempre que a soma acima convergir absolutamente.

Se $X \in Y$ forem independentes, então

$$\mathbb{E}[XY] = \sum_{x,y} xy \ \mathbb{P}_r\Big(X = x, \ Y = y\Big) = \mathbb{E}(X) \ \mathbb{E}(Y). \tag{3-13}$$

Novamente, a ideia se estende para uma família (vetores) de v.a. $\mathbf{X} = (X_1, X_2, X_3, ... X_n)$. Se \mathbf{X} for composto por v.a. independentes, então,

$$p_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}_r(X_1 = x_1)\mathbb{P}_r(X_2 = x_2)...\mathbb{P}_r(X_n = x_n),$$
(3-14)

е

$$\mathbb{E}(X_1 X_2 \dots X_n) = \mathbb{E}(X_1) \mathbb{E}(X_2) \dots \mathbb{E}(X_n).$$
(3-15)

3.4 S

Soma de Variáveis Aleatórias

Dados X e Y v.a. discretas independentes com função de massa conjunta definida, e Z = X + Y. Se os valores assumidos por Z satisfizerem z = x + y, onde x são os valores assumidos por X e y os valores assumidos por Y, quando X = x, então

$$\mathbb{P}_r(Z=z) = \sum_{x,y} \mathbb{P}_r(X=x, Y=y) \text{ para } z \in \mathbb{R}$$
$$\mathbb{P}_r(Z=z) = \sum_x \mathbb{P}_r(X=x, Y=z-x). \tag{3-16}$$

De acordo com a fórmula da convolução, demonstrado por Grimmet [19] na página 45, se X e Y são duas variáveis aleatórias independentes em

 $(\Omega, \mathcal{F}, \mathbb{P}_r)$, então a função de massa de Z = X + Y é dada por:

$$\mathbb{P}_r(Z=z) = \sum_x \mathbb{P}_r(X=x, \ Y=z-x) = \sum_x \mathbb{P}_r(X=x)\mathbb{P}_r(Y=z-x).$$
(3-17)

3.5 Processos Estocásticos

Muitos fenômenos do mundo real envolvem a variação de um parâmetro, por exemplo tempo. Alguns exemplos são fenômenos meteorológicos, índices do mercado de ações, dados de poluição do ar, dados demográficos e pesquisas de acompanhamento político [27]. Para modelar probabilisticamente esse tipo de fenômeno, em geral utiliza-se um tipo de objeto probabilístico chamado de processo estocástico.

Dado um conjunto de variáveis aleatórias RV, do inglês random variable, em $(\Omega, \mathcal{F}, \mathbb{P}_r)$. Um processo estocástico \mathcal{X} é uma função definida por:

$$\mathcal{X}: T \to RV(\Omega, \mathcal{F}, \mathbb{P}_r)$$
$$t \mapsto \mathcal{X}(t),$$

tal que T pode ser um conjunto de parâmeros discretos ou contínuos.

Assim, para um valor específico de $t \in T$, $\mathcal{X}(t)$ é uma variável aleatória definida em $(\Omega, \mathcal{F}, \mathbb{P}_r)$. O conjunto de índices t não descreve, necessariamente, o tempo. Esse parâmetro também pode ser usado para descrever uma localização espacial.

Se T for um conjunto de parâmetros finitos, isto é, $T = \{t_1, t_2, ..., t_m\}$, o processo estocástico em questão pode ser entendido como um vetor aleatório.

Um processo estocástico pode ser discreto ou contínuo, estacionário ou não. Alguns dos processos mais conhecidos são: passeios aleatórios, processos de Markov, processos de Lévy, processos Gaussianos e processo de ramificação. Dentre esses processos, será utilizado neste trabalho o processo de ramificação.

3.6 Método de Monte Carlo

Definidos alguns dos conceitos básicos de probabilidade, vamos agora a parte computacional. Para se obter estatísticas amostrais da resposta de um sistema que depende de parâmetros estocásticos, modelados como variáveis aleatórias discretas, seguimos uma determinada organização [36]:

- constrói-se um modelo determinístico para o sistema,
- constrói-se um modelo estocástico, determinando funções massa para os parâmetros considerados aleatórios no sistema,

 aplica-se o método de Monte Carlo para obtermos inferências estatísticas sobre a resposta do sistema.

De forma esquematizada temos:



Figura 3.1: Esquematização de como obter estatísticas amostrais da resposta de um sistema que depende de parâmetros estocásticos, modelados como variáveis aleatórias discretas.

Assim, método de Monte Carlo é uma classe de algoritmos computacionais que dependem de uma amostragem aleatória para obter resultados numéricos. Nele, amostras são geradas e, posteriormente, processadas através do modelo determinístico de forma a obter amostras da resposta do sistema. Com as amostras da resposta do sistema constrói-se um modelo estatístico da saída do sistema, ou seja calculam-se estatísticas amostrais, como por exemplo, média, variância e histogramas normalizados.

Para avaliar as estatísticas amostrais, definimos um erro no inicio do processo e analisamos se elas estão dentro dessa margem, baseados no teorema do limite central e na lei dos grandes números. A seguir, estão descritos esses dois teoremas.

3.6.1 Lei dos Grandes Números

A lei dos grandes números é dada da seguinte maneira: considerando uma sequência $X_1, X_2, X_3, ..., X_n$ de v.a. independentes e identicamente distribuídas, com média μ_X e variância σ_X , e a v.a. $S_n = X_1 + X_2 + X_3 + ... + X_n$. Temos que S_n/n converge em média quadrática (média-r, com r = 2) para μ_X . Isto é,

$$\left\|\frac{S_n}{n} - \mu_X\right\| = \sqrt{\mathbb{E}\left[\left(\frac{S_n}{n} - \mu\right)^2\right]} \to 0, \text{ quando } n \to +\infty.$$
(3-18)

Assim,

$$\mu_{S_n/n} = \mathbb{E}\left[\frac{S_n}{n}\right] = \frac{1}{n} (\mathbb{E}[X_1] + \mathbb{E}[X_2] + \dots + \mathbb{E}[X_n]) = \mu_x, \qquad (3-19)$$

е

$$\sigma_{S_n/n}^2 = \mathbb{E}\left[\left(\frac{S_n}{n} - \mu\right)^2\right] = \frac{1}{n^2} (var[X_1] + var[X_2] + \dots + var[X_n]) = \frac{\sigma_X^2}{n}.$$
 (3-20)

Com isso, observamos que conforme aumentamos o número n a variância de S_n/n decai com uma razão de 1/n.

3.6.2 Teorema do Limite Central

Quanto ao teorema do limite central temos em [32]: Seja $X_1, X_2, X_3, ..., X_n$ uma sequência de v.a. independentes e identicamente distribuídas, com média μ e variância σ^2 . Dado $S_n = X_1 + X_2 + X_3 + ... + X_n$ e $Z_n = \frac{S_n - \mu_n}{\sigma \sqrt{n}}$. Quando $n \to \infty$, Z_n converge em distribuição (lei) para uma variável aleatória Z normal de média nula e variância um. Assim,

 $\mathbb{P}_{Z_n}(Z_n \leq x) \to \mathbb{P}_Z(Z \leq x)$ quando $n \to +\infty, \forall x$ onde \mathbb{P}_Z é contínua. (3-21)

3.7 Quantificação e Propagação de Incertezas

A quantificação de incertezas se faz presente em sistemas onde há parâmetros modelados por objetos aleatórios. As incertezas associadas a esses parâmetros são propagadas para a resposta do sistema e podem variar ao longo do tempo, caso a resposta do sistema dependa do tempo. Em geral, a resposta de um sistema com parâmetros estocásticos depende diretamente dos modelos estocásticos atribuídos a esses parâmetros [23, 24].

Para quantificarmos a incerteza presente na resposta do sistema, precisamos caracterizar essa resposta de uma forma probabilística. Se a resposta for, por exemplo, um processo estocástico, em geral, deseja-se caracterizar esse processo através de suas distribuições de probabilidade marginais, ou seja, distribuições de seções do processo. Uma forma de se obter aproximações para essas distribuições é construir histogramas normalizados através de inferências estatísticas das resposta conforme o Método de Monte Carlo [32].

4 Função Geradora de Probabilidade

"Uma função geradora é um dispositivo um tanto semelhante a uma bolsa. Em vez de carregar muitos pequenos objetos separados,[...], colocamos todos em uma bolsa, e então temos apenas um objeto para carregar, a bolsa.", George Pólya [30].

4.1 Função Geradora

A função geradora de uma sequência é definida pela soma de uma série de potência

$$u_0 + u_1 s + u_2 s^2 + u_3 s^3 + \dots, (4-1)$$

na qual u_0, u_1, u_2 , etc. são os valores que compõem a sequência.

A sequência $1, 2, 3, 4, \ldots$, por exemplo, tem como função geradora a equação dada por:

$$G(s) = 1 + 2s + 3s^{2} + 4s^{3} \dots = \frac{1}{(1-s)^{2}} , \qquad (4-2)$$

para valores de $s \in (0, 1)$ tais que a soma convirja.

Para obter o n-ésimo valor que compõe a sequência, derivamos a equação 4-2 n vezes, dividimos por n! e avaliamos quando s = 0 da seguinte forma:

$$\begin{aligned} u_0 &= G(s=0) = \left. \frac{1}{(1-s)^2} \right|_{s=0} = \left. \frac{1}{(1-0)^2} = 1; \\ u_1 &= G'(s=0) = \left. \frac{2}{1!(1-s)^3} \right|_{s=0} = \left. \frac{2}{1!(1-0)^3} = 2; \\ u_2 &= G''(s=0) = \left. \frac{6}{2!(1-s)^4} \right|_{s=0} = \left. \frac{6}{2!(1-0)^4} = 3; \\ u_3 &= G'''(s=0) = \left. \frac{24}{3!(1-s)^5} \right|_{s=0} = \left. \frac{24}{3!(1-0)^5} = 4; \\ \vdots \end{aligned}$$

Para a sequência 1, 2, 4, 8, ... temos:

$$G(s) = 1 + 2s + 4s^2 + 8s^3 + \dots = \sum_{n=0}^{\infty} (2s)^n = \frac{1}{1 - 2s} , \qquad (4-3)$$

para valores de $s \in (0, 1)$ tais que a soma convirja. Para recuperar o n-ésimo termo da sequência, derivamos n vezes a equação 4-3 dividimo-a por n! e avaliamo-na em s = 0.

$$u_{0} = G(s = 0) = \frac{1}{(1 - 2s)}\Big|_{s=0} = \frac{1}{(1 - 0)} = 1;$$

$$u_{1} = G'(s = 0) = \frac{2}{1!(1 - 2s)^{2}}\Big|_{s=0} = \frac{2}{1!(1 - 0)^{2}} = 2;$$

$$u_{2} = G''(s = 0) = \frac{8}{2!(1 - 2s)^{3}}\Big|_{s=0} = \frac{8}{2!(1 - 0)^{3}} = 4;$$

$$u_{3} = G'''(s = 0) = \frac{48}{3!(1 - 2s)^{4}}\Big|_{s=0} = \frac{48}{3!(1 - 0)^{4}} = 8;$$

$$\vdots$$

4.2 Função Geradora de Probabilidade - PGF

Dada uma variável aleatória discreta X, no espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P}_r)$, com probabilidades p_0, p_1, p_2, \ldots . A função geradora de probabilidade (PGF) de X é definida por

$$G_X(s) = p_0 + p_1 s + p_2 s^2 + \dots, \qquad (4-4)$$

para todo valor de s que a sequência converge. Isto é, a função dada por 4-4 gera a sequência de valores p_0 , p_1 , p_2 , etc.

Analisando a função $G_X(s)$ em s = 1 temos

$$G_X(s=1) = p_0 + p_1 1 + p_2 1^2 + \dots = \sum_{n=0}^{\infty} p_n = 1$$

$$G_X(1) = 1,$$
(4-5)

ou seja, o somatório de todas as probabilidades é igual a 1.

Se analisarmos agora a condição de s = 0, teremos a probabilidade p_0 de X assumir o valor 0:

$$G_X(s=0) = p_0 + p_1 0 + p_2 0^2 + \dots$$

$$G_X(0) = p_0.$$
(4-6)

Variáveis aleatórias discretas que apresentam a mesma função geradora de probabilidade tem também a mesma função de massa. Sendo assim
$$G_X(s) = G_Y(s) \text{ se, e somente se,}$$

$$\mathbb{P}_r(X = k) = \mathbb{P}_{\searrow}(Y = k), \text{ para } k=0,1,2,3,\dots$$
(4-7)

A seguir, estão descritas as funções geradoras de probabilidade das variáveis aleatórias que serão usadas neste trabalho [19]:

Sendo X uma v.a. Bernoulli com parâmetro p:

$$G_X(s) = q + ps. \tag{4-8}$$

Sendo X uma v.a. Binomial com parâmetros $p \in m$:

$$G_X(s) = (q + ps)^m.$$
 (4-9)

Sendo X uma v.a. Geométrica - $1 \operatorname{com} parâmetro p$:

$$G_X(s) = \frac{ps}{1 - qs}.\tag{4-10}$$

Sendo X uma v.a. Geométrica - $0 \operatorname{com} parâmetro p$:

$$G_X(s) = \frac{p}{1 - qs}.\tag{4-11}$$

Nas quatro PGF descritas acima, o valor de q é dado por q = 1 - p. Um exemplo de uma geométrica-0 e uma geométrica-1 serão dados nos capítulos a seguir.

4.3

Cálculo de Média e Variância de Variáveis Aleatórias Através da Função Geradora de Probabilidade

De acordo com o teorema 4.23 de [19]: seja X uma variável aleatória com sua função geradora de probabilidade $G_X(s)$, então a n-ésima derivada (representada por $G^{(n)}$) de $G_X(s)$ avaliado em s = 1 é dado por:

$$G_X^{(n)}(s=1) = \mathbb{E}(X[X-1]...[X-n+1]).$$
(4-12)

Desta definição temos que o valor esperado de X é dado por:

$$\mathbb{E}(X) = G'_X(s=1), \tag{4-13}$$

e a variância de X é dada por:

$$var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$
$$var(X) = (G''_X(1) + G'_X(1)) - (G_X(1)^2).$$
(4-14)

4.4 Soma de Varáveis Aleatórias Independentes

Quando se trabalha com soma de variáveis aleatórias independentes a PGF se torna uma ferramenta muito útil, já que ela elimina a necessidade de se trabalhar com convolução, (soma de v.a., como mostrada na seção 3.4). Se duas variáveis aleatórias discretas $X \in Y$ são independentes então a soma delas tem PGF igual a

$$G_{X+Y}(s) = G_X(s)G_Y(s).$$
 (4-15)

Como prova desta equação:

$$G_{X+Y}(s) = \mathbb{E}(s^{X+Y}) = \mathbb{E}(s^X s^Y)$$

= $\mathbb{E}(s^X)\mathbb{E}(s^Y)$
= $G_X(s)G_Y(s).$ (4-16)

Analogamente, para a soma $S_n = X_1 + X_2 + X_3 + ... + X_n$ de *n* variáveis aleatórias independentes entre si, a função geradora de $S_n \in G_{S_n}(s) = G_{X_1}(s)G_{X_2}(s)...G_{X_n}(s).$

Derivado deste conceito, podemos definir a soma randômica. Seja X_1 , X_2 , X_3 , ... variáveis aleatórias discretas IID (independentes e identicamente distribuídas), isto é, independentes entre si e com a mesma PGF, e N uma outra variável aleatória discreta. Então, a soma dessas variáveis é dada por

$$Z = X_1 + X_2 + \dots + X_N,$$

e a função geradora de probabilidade de Z é

$$G_Z(s) = G_N(G_X(s)),$$
 (4-17)

onde $G_X(s)$ é a função geradora de $X_1, X_2, ..., X_N$, identicamente distribuídas.

Como prova temos:

$$G_Z(s) = \mathbb{E}(s^Z) = \mathbb{E}(s^{X_1 + X_2 + \dots + X_N})$$

$$= \sum_{n=0}^{\infty} \mathbb{E}(s^{X_1 + X_2 + \dots + X_N} | N = n) \mathbb{P}_r(N = n), \text{ pelo Teorema da Partição}$$

$$= \sum_{n=0}^{\infty} \mathbb{E}(s^{X_1 + X_2 + \dots + X_n}) \mathbb{P}_r(N = n),$$

$$= \sum_{n=0}^{\infty} G_X(s)^n \mathbb{P}_r(N = n), \text{ pela equação 4-15}$$

$$= G_N(G_X(s)).$$
(4-18)

Calculando $G^\prime{}_Z(s)$ aplicado em s=1 temos:

$$G'_{Z}(s) = G'_{N}(G_{X}(x))G'_{X}(x)$$

$$G'_{Z}(1) = G'_{N}(G_{X}(1))G'_{X}(1) = G'_{N}(1)G'_{X}(1)$$

$$\mathbb{E}(Z) = \mathbb{E}(N)\mathbb{E}(X),$$
(4-19)

onde $\mathbb{E}(X)$ é a média da variável aleatória X e $\mathbb{E}(N)$ é a média da variável aleatória N.

5 Construção do Modelo Probabilístico para o Contágio e a Propagação de uma Doença

Neste capítulo vamos aplicar os conceitos vistos até aqui no cenário de propagação de uma doença. Analisaremos o contágio e a propagação com uma abordagem estocástica.

5.1 Construção do Modelo Probabilístico para o Contágio e a Propagação de uma Doença

Esta seção trata da modelagem probabilística do contágio entre indivíduos da população e da propagação da doença [5].

O número de indivíduos contaminados por cada membro já infectado j da população é modelado como uma variável aleatória (v.a.) discreta denominada B_j . Supõe-se que as v.a. B_j sejam independentes e identicamente distribuídas (IID). Com isso, denominamos B a variável aleatória que modela o contágio.

O número de infectados ao longo do tempo é modelado por um processo estocástico I de parâmetro discreto chamado de processo de ramificação. Dado um conjunto de parâmetros discretos $N = \{0, 1, 2, 3, ...\}$, representando gerações ao longo do tempo, para cada $n \in N$, $I(n) = I_n$ é uma variável aleatória discreta. I_n representa o número de infectados na n-ésima geração e é dado pela soma do número de pessoas que foram contagiadas por cada um dos infectados na geração anterior n - 1.

A Figura 5.1 ilustra as cinco primeiras gerações de uma realização do processo de ramificação I. Nesta realização, o processo de contaminação começa com um indivíduo infectado, $i_0 = 1$. Na geração n = 1, aparecem 3 indivíduos contaminados pelo indivíduo infectado inicialmente, $i_1 = 3$. Esse número de infectados é obtido fazendo-se uma realização da variável aleatória B. Na próxima geração, n = 2, o número total de infectados é a soma do número de pessoas que foram contagiadas por cada um dos três infectados na geração anterior. Para obter esse número, geramos três realizações independentes de B e somamos os valores sorteados. No exemplo ilustrado, o primeiro infectado contaminou apenas um indivíduo e os outros dois não fizeram nenhuma contaminação. Dessa forma $i_2 = 1$. Na próxima



Figura 5.1: Uma realização do processo de ramificação que modela o número de infectados ao longo de geraç o
es. Considera-se $i_0=1$

geração, n = 3, o único indivíduo contaminado na geração anterior contaminou quatro novos indivíduos, fazendo $i_3 = 4$. Na geração n = 4, o número total de infectados, que é a soma do número de pessoas que foram contagiadas por cada um dos quatro infectados em n = 3, e foi obtido gerando-se quatro realizações independentes de B e somando os valores sorteados. No exemplo ilustrado, o primeiro infectado contaminou dois indivíduos, o segundo infectado não fez nenhuma contaminação e o terceiro e quarto fizeram duas contaminações cada. Totalizando $i_4 = 7$. Se nenhum desses sete indivíduos contaminados fizer novas contaminações, teremos $i_n = 0$, $\forall n \geq 5$, e nesse caso a doença estará extinta. A probabilidade de extinção de uma doença é uma das variáveis de interesse desse trabalho.

5.2 Função Geradora de Probabilidade da Propagação de uma Doença

Utilizando as definição estabelecidas no capítulo 4 e as variáveis aleatórias definidas na seção 5.1 temos:

- $-I(n) = I_n$ número total de infectados em uma geração;
- $-B_j =$ número de indivíduos contaminados pelo j-ésimo indivíduo da geração anterior.

Assim,

$$I_{0} = 1$$

$$I_{1} = B_{I_{0}}$$

$$I_{2} = B_{1} + B_{2} + \dots + B_{I_{1}}$$

$$\vdots$$

$$I_{n} = B_{1} + B_{2} + \dots + B_{I_{n-1}} = \sum_{j=0}^{I_{n-1}} B_{j}, \text{ para } n > 0, \quad (5-1)$$

tal que B_j são os novos contagiados pelos infectados da geração n-1.

Como vimos na seção 4.4, se $I_n = B_1 + B_2 + \ldots + B_{I_{n-1}}$ é a soma randômica de variáveis aleatórias IID, então:

$$G_{I_n}(s) = G_{I_{n-1}}(G_B(s)).$$
(5-2)

Vale, também, destacar que:

$$G_{I_n}(s) = G_{I_{n-1}}(G_B(s))$$

= $G_{I_{n-2}}(G_B(G_B(s)))$
= $G_{I_{n-3}}(G_B(G_B(G_B(s))))$
:
= $G_B(G_B(G_B(...(s)...)))$
= $G_B(G_{I_{n-1}}(s)),$ (5-3)

na qual a função geradora de probabilidade do contágio $G_B(s)$ é descrita por

$$G_B(s) = \sum_{k=0}^{\infty} p_k s^k, \tag{5-4}$$

onde p_k é a probabilidade de B assumir o valor k.

5.3 Função de Distribuição de Probabilidade do Total de Infectados

Dada uma variável aleatória discreta X, no espaço de probabilidade $(\Omega, \mathcal{F}, \mathbb{P})$, com probabilidades p_0, p_1, p_2, \ldots , a função geradora de probabilidade de X é definida por

$$G_X(s) = p_0 + p_1 s + p_2 s^2 + \dots ,$$
 (5-5)

para todo valor de s que a sequência converge. Isto é, a função dada por 5-5 gera a sequência de valores p_0, p_1, p_2, \ldots Para recuperarmos o valor do n-ésimo termo da sequência, da derivamos n vezes a equação de $G_X(s)$, dividimos ela

por n! e avaliamo-na em s = 0.

O mesmo conceito se aplica a PGF da propagação de uma doença (equação 5-2). Assim é possível obter os valores de p_0, p_1, p_2, \ldots em cada uma das n gerações, onde $p_{n_0}, p_{n_1}, p_{n_2}, \ldots$ são as probabilidades de se obter $0, 1, 2, \ldots$ infectados na n-ésima geração.

Isto é, se

$$G_{I_1}(s) = G_{I_0}(G_B(s)) = G_B(s)$$
(5-6)

então

$$G_{I_1}(s=0) = p_{1_0};$$

$$G'_{I_1}(s=0) = p_{1_1};$$

$$G''_{I_1}(s=0) = p_{1_2};$$

$$G'''_{I_1}(s=0) = p_{1_3};$$

$$G''''_{I_1}(s=0) = p_{1_4}.$$
(5-7)

Para a n-ésima geração teremos:

$$G_{I_n}(s) = G_{I_{n-1}}(G_B(s)), (5-8)$$

 $ent \tilde{a} o$

$$G_{I_n}(s = 0) = p_{n_0};$$

$$G'_{I_n}(s = 0) = p_{n_1};$$

$$G''_{I_n}(s = 0) = p_{n_2};$$

$$G'''_{I_n}(s = 0) = p_{n_3};$$

$$G''''_{I_n}(s = 0) = p_{n_4};$$

$$\vdots$$

$$G_{I_n}^{(y)}(s = 0) = p_{n_y}.$$
(5-9)

5.4 Algumas Estatísticas da Propagação

5.4.1 Média e Variância do Total de Infectados

A seguir estão definidas algumas das estatísticas do processo que modela a propagação [1, 21].

Dada a equação 4-13, fazendo $X = I_n$, temos:

$$\mathbb{E}(I_n) = G'_{I_n}(1)
= G'_{I_{n-1}}(G_B(1))G'_B(1)
= G'_{I_{n-1}}(1)G'_B(1)
= \mathbb{E}(I_{n-1}) \mu_B
= (\mathbb{E}(I_{n-2})\mu_B)\mu_B
\vdots
\mathbb{E}(I_n) = \mu_B^n,$$
(5-10)

onde $\mu_B = \mathbb{E}(B)$, valor esperado do contágio B.

Seja $\sigma^2(I_n)$ a variância do número de infectados da geração n. Utilizando o mesmo raciocínio da média para definir a variância do número de infectados em cada geração n, temos:

$$var(I_{n}) = [(G_{I_{n}}''(1)] + [G_{I_{n}}'(1))] - [(G_{I_{n}}(1)^{2})]$$

$$= [G_{I_{n}}''(1)] + [G_{B}'(G_{I_{n-1}}(1))G_{I_{n-1}}'(1)] - [(G_{I_{n}}(1)^{2})]$$

$$= [G_{B}''(G_{I_{n-1}}(1))G_{I_{n-1}}'(1) + G_{I_{n-1}}''(1)G_{B}'(G_{I_{n-1}}(1))] + [G_{B}'(G_{I_{n-1}}(1))G_{I_{n-1}}'(1)] - [(G_{I_{n}}(1)^{2})]$$

$$= [G_{B}''(1)G_{I_{n-1}}'(1) + G_{I_{n-1}}''(1)G_{B}'(1)] + [G_{B}'(1)G_{I_{n-1}}'(1)] - [1]$$

$$= [(G_{B}''(1)G_{I_{n-1}}'(1) + G_{I_{n-1}}''(1)\mu_{B})] + [\mu_{B}G_{I_{n-1}}'(1)] - [1].$$
(5-11)

Ainda, de acordo com 4-14

se $\sigma^{2} = (G''_{B}(1) + G'_{B}(1)) - (G_{B}(1)^{2})$ $\sigma^{2} = (G''_{B}(1) + \mu_{B}) - (1),$ então

$$\sigma_{I_n}^2 = (G_{I_n}''(1) + \mathbb{E}(I_n)) - (1), \qquad (5-12)$$

$$\sigma_{I_{n-1}}^2 = (G_{I_{n-1}}''(1) + \mathbb{E}(I_{n-1})) - (1).$$
(5-13)

Sabendo que $\mathbb{E}(I_n) = \mu_B^n$ e $\mathbb{E}(I_{n-1}) = \mu_B^{n-1}$ e subtituindo 5-12 e 5-13 em 5-11 temos:

$$\sigma_{I_n}^2 = \mu_B^{n-1} \sigma^2 (1 + \mu_B + \mu_B^2 + \mu_B^3 + \dots + \mu_B^{n-1}).$$
 (5-14)

Com isso:

$$\sigma_{I_n}^2 = \begin{cases} n\sigma^2 & \text{se } \mu_B = 1\\ \sigma^2 \mu_B^{n-1} \left(\frac{\mu_B^n - 1}{\mu_B - 1}\right) & \text{se } \mu_B \neq 1, \end{cases}$$
(5-15)

onde μ_B é a média de B, e σ^2 a variância da mesma.

Uma análise interessante a partir dessas equações é a influência do modelo escolhido para o contágio no comportamento da epidemia a longo prazo. Se analisarmos $\lim_{n\to\infty} \mathbb{E}(I_n) \in \lim_{n\to\infty} var(I_n)$, em 5-10 e 5-15 respectivamente, temos:

$$\lim_{n \to \infty} E(I_n) \to \begin{cases} 0 \text{ se } \mu_B < 1 \\ 1 \text{ se } \mu_B = 1 \\ \infty \text{ se } \mu_B > 1 \end{cases}, \qquad \lim_{n \to \infty} var(I_n) \to \begin{cases} 0 \text{ se } \mu_B < 1 \\ \infty \text{ se } \mu_B = 1 \\ \infty \text{ se } \mu_B > 1 \end{cases}$$
(5-16)

A partir da equação acima podemos verificar que:

- se μ_B < 1, a média de infectados quando t → ∞ tende a zero. Isso significa que a longo prazo essa epidemia vai se extinguir com certeza;
- se $\mu_B = 1$, a média de infectados flutua em torno de um. Este é um caso crítico, uma vez que o tamanho médio do total de infectados é igual a 1 em todas as *n* gerações. Neste caso, flutuações aleatórias garantem que o total de infectados assumirá o valor 0 mais cedo ou mais tarde. Daí em diante teremos a extinção da doença. Quando olhamos para a variância, vemos que ela tende a infinito;
- e se $\mu_B > 1$, temos que a média e a variância tentem a infinito. A probabilidade da epidemia se extinguir neste cenário é diferente de 100%.

5.5 Probabilidade de Extinção

Na seção anterior vimos que se $\mu_B \leq 1$ a epidemia se extingue com probabilidade de 100%. Mas o que acontece se $\mu_B > 1$? Para determinar a probabilidade de extinção da doença definiremos 2 eventos.

Primeiro definimos um evento A: extinção da doença na n-ésima geração. Isto é, $I_n = 0$. Utilizando a equação 5-3 e avaliando ela em s = 0 temos:

$$G_{I_n}(s) = G_B(G_{I_{n-1}}(s))$$

$$G_{I_n}(0) = G_B(G_{I_{n-1}}(0)) = e_n.$$
(5-17)

Assim, chamamos de e_n a probabilidade de A.

Depois, um segundo evento, B: extinção da doença. No qual, de acordo com o teorema 1.54 de [19], a probabilidade deste evento ocorrer é dado por: $\lim_{n\to\infty} e_n$, representado por e.

Assim:

$$e_n = G_{I_n}(0) = G_B(G_{I_{n-1}}(0)) \quad \rightarrow \quad e_n = G_B(e_{n-1})$$
$$\lim_{n \to \infty} e_n = e \quad \rightarrow \quad e = G_B(e), \tag{5-18}$$

onde a menor raiz da equação $e = G_B(e)$ no intervalo [0, 1] é a probabilidade da epidemia se extinguir.

Nos próximos capítulos serão feitas simulações baseadas no modelo e nas análises propostas neste capítulo.

Influência do Modelo do Contágio no Comportamento da Doença

Este capítulo traz as comparações entre os diferentes modelos escolhidos para modelar o contágio da doença, além do contexto em que elas estão inseridas e suas consequências.

6.1

6

Estratégia para Analisar a Influência do Modelo Probabilístico do Contágio na Propagação da Doença

Para analisar a influência do modelo probabilístico da v.a. que modela o contágio entre indivíduos na propagação da doença e probabilidade de extinção, foram feitas simulações de Monte Carlo e construídos modelos estatísticos para a propagação da doença considerando-se 3 famílias diferentes de v.a. B: binomial, geométrica-1 e geométrica-0 [6]. Para cada família, 21 distribuições diferentes foram selecionadas e, para cada distribuição, 4.000 simulações do processo de ramificação da geração n = 0 a n = 20 foram computadas. No total, foram realizadas 252 realizações do processo estocástico de ramificação e aproximadamente 5 milhões de amostras da variável aleatória do contágio B. Para fazer todas as simulações gastou-se aproximadamente 30 horas de processamento (tempo de CPU). O elevado custo computacional em tempo e volume de dados, caracteriza o problema como *big data*. Um estudo de convergência foi desenvolvido para determinar o número de simulações [36]. Uma esquematização dos parâmetros selecionados é mostrado na figura 6.1.

Para analisar o comportamento estocástico da epidemia ao longo das gerações, foram computadas estatísticas amostrais, como média, variância e probabilidade de extinção e histogramas normalizados do número de indivíduos infectados para as diferentes gerações do processo de ramificação. Com esses histogramas é possível observar a propagação de incertezas [32] na disseminação da doença. Dentre todas as simulações realizadas, os resultados mais expressivos obtidos com as simulações de Monte Carlo são apresentados nas seções a seguir.



Média, variância e histogramas normalizados

Figura 6.1: Estratégia para analisar a influência do modelo probabilístico do contágio na propagação da doença.

6.2 Modelo Probabilístico de ${f B}$ - Binomial (m, p)

O primeiro modelo probabilístico adotado foi a binomial(m, p).

A distribuição binomial pode ser interpretada, junto com os conceitos de epidemiologia, como um experimento no qual um indivíduo encontra com um número m de pessoas, com uma probabilidade p de infectar cada uma delas. Depois desse encontro, avaliamos se as m pessoas foram ou não infectadas, contando o número de novos contagiados. Esse cenário retrata uma sociedade com regras de distanciamento restritas, onde, cada indivíduo tem um contato limitado a, no máximo, m indivíduos. Nas simulações, fazemos m = 3.

Com 4000 realizações do processo de contágio foram construídos histogramas normalizados, com o objetivo de analisar o comportamento da epidemia ao longo das gerações. Para ilustrar o tipo de resultado obtido com as simulações, mostramos nos gráficos da figura 6.2 os histogramas normalizados do número de infectados em diferentes gerações da doença, com o contágio modelado pela binomial(3, 0.5).

Os histogramas da figura 6.2 mostram, em particular, a propagação de incerteza inerente ao sistema. É possível observar a mudança da distribuição de probabilidade conforme as gerações de infectados vão se desenvolvendo.

Como vimos no capítulo 4 uma distribuição binomial tem media $\mu_B = mp$ e sua função geradora de probabilidade é $G_B(s) = (q + ps)^m$.



Histogramas com B = binomial (3, 0.50)

Figura 6.2: Histogramas normalizados com o total de infectados nas gerações 2, 4, 6 e 21, construídos com 4000 realizações.

Utilizando o desenvolvimento proposto na seção 5.4.1 concluímos que, para um valor de m pré-estabelecido, temos:

Uma outra análise a ser feita é determinar a probabilidade e de extinção da doença. Com 5-18 temos

$$G_B(e) = (q + pe)^m = e.$$
 (6-2)

Dentre todos os resultados obtidos, destacamos a seguir os resultados para a propagação da doença e probabilidade de extinção para 3 diferentes valores de media μ_B . Serão analisados o caso de $\mu_B < 1$, $\mu_B = 0$ e $\mu_B = 1$.

6.2.1 Contágio B = binomial (3, 0.16), com $\mu_B < 1$



Histogramas e Funções de Massa de I_n com B = binomial (3, 0.16)

Figura 6.3: Comparação entre os histogramas normalizados construídos com 4000 realizações e as funções de massa calculadas analiticamente.

A figura 6.3 mostra os histogramas normalizados e as funções de massa do número de infectados em diferentes gerações com o do contágio modelado por B = binomial (3, 0.16). É possível observar em vermelho que os valores obtidos analiticamente de acordo com 5-9 batem com os valores obtidos através das 4000 realizações, mostrando que, o algorítimo e a quantidade de realizações estão adequados.



Figura 6.4: Contágio B = binomial (3, 0.16), com $\mu_B < 1$.

A figura 6.4 mostra a média e variância da propagação da doença e a probabilidade de extinção quando o contágio é modelado por uma

binomial(3, 0.16). O contágio de um indivíduo infectado é limitado a três novos indivíduos. Como, neste caso, $\mu_B < 1$ vemos que a média e a variância tendem a zero (equação 5-16 e figura 6.4a), indicando que, eventualmente, a epidemia irá se extinguir com 100% de probabilidade (figura 6.4b).

6.2.2 Contágio ${f B}=$ binomial $({f 3},\ {f 0}.{f 3}{f 3})$, com $\mu_{f B}={f 1}$

Histogramas e Funções de Massa de $I_n \text{ com } B = \text{binomial } (3, 0.33)$



Figura 6.5: Comparação entre os histogramas normalizados construídos com 4000 realizações e as funções de massa calculadas analiticamente.

A figura 6.5 mostra os histogramas normalizados e as funções de massa do número de infectados em diferentes gerações com o do contágio modelado por B = binomial (3, 0.33). Nela é possível observar a propagação da incerteza ao longo das gerações, e a adequação dos valores calculados ao modelo proposto.

A figura 6.6 mostra a média e variância do número de infectados ao longo das gerações e probabilidade de extinção com o contágio modelado por uma binomial(3,0.33). Como, neste caso, $\mu_B = 1$ vemos que a média permanece em torno de 1 enquanto a variância cresce linearmente conforme as gerações vão passando (equação 5-16 e figura 6.6a), indicando que, eventualmente, a epidemia irá se extinguir com 100% de probabilidade (figura 6.6b).



Figura 6.6: Contágio B = binomial (3, 0.33), com $\mu_B = 1$.

6.2.3 Contágio B= binomial $(3,\ 0.50)$, com $\mu_{\rm B}>1$

A figura 6.7 mostra os histogramas normalizados e as funções de massa do número de infectados em diferentes gerações com o contágio modelado por B = binomial (3, 0.50). Vale destacar nela a queda no valor da probabilidade de se ter apenas 1 infectado nas gerações 4 e 5.

Histogramas e Funções de Massa de I_n com B = binomial (3, 0.50)



Figura 6.7: Comparação entre os histogramas normalizados construídos com 4000 realizações e as funções de massa calculadas analiticamente.

A figura 6.8 mostra o caso da média e variância do número de infectados ao longo das gerações e probabilidade de extinção quando o contágio é modelado por uma binomial(3,0.33). Assim como nos casos anteriores, o contágio continua limitado à 3 pessoas. No entanto, como neste caso, $\mu_B > 1$ vemos que a média e a variância crescem conforme as gerações vão passando,



Figura 6.8: Contágio B = binomial (3, 0.50), com $\mu_B > 1$.

tendendo a infinito (equação 5-16 e figura 6.8a). Com base no equacionamento proposto em 5-18 temos, então, uma probabilidade de extinção $e \neq 1$. Para os parâmetros em questão, a probabilidade é de e = 23,6% (figura 6.8b).

Como é possível ver pelos gráficos desta seção, os resultados das simulações numéricas estão de acordo com a resolução desenvolvida a partir dos conceitos de probabilidade e estatísticas apresentados. As duas próximas seções mostram os resultados do trabalho quando adotamos modelos probabilísticos diferentes para o contágio B.

6.3 Modelo Probabilístico de ${ m B}$ - Geométrica Parametrizada(p)

O segundo modelo analisado foi uma parametrização da geométrica(p), chamada neste trabalho de geométrica - 0. Nesta análise um indivíduo infectado que sai de casa não tem um limite máximo de infectados. Ele pode sair também e não infectar nenhum novo indivíduo. Esse modelo retrata uma sociedade com regras de distanciamento social baixas, com contato entre pessoas mais alto e menos restrito, mas no qual algumas pessoas praticam o isolamento social e respeitam regras de distanciamento, apesar de saírem ainda na rua.

O histograma da figura 6.9 mostra a propagação de incerteza no modelo utilizado.

A função geradora de probabilidade do contágio é dada por $G_B(s) = \frac{p}{1-qs}$, e sua média é $\mu_B = \frac{1-p}{p}$. Como é possível indivíduos saírem de suas casas e não contaminarem ninguém, existe a possibilidade da doença se extinguir da sociedade. Então,



Histogramas com B = geométrica-0 (0.6)

Figura 6.9: Histogramas normalizados com o total de infectados nas gerações 2, 4, 6 e 12, construídos com 4000 realizações.

$$\begin{cases} p > 1 - p \quad \therefore \ \mu_B < 1 \ \to \ \text{epidemia se extingue;} \\ \\ p = 1 - p \quad \therefore \ \mu_B = 1 \ \to \ \text{epidemia se extingue;} \\ \\ p < 1 - p \quad \therefore \ \mu_B > 1 \ \to \ \text{epidemia pode, ou não se extinguir, com probabilidade} \ e = \frac{p}{1 - p}. \end{cases}$$

6.3.1 Contágio ${f B}=$ geométrica - 0 (0.4), com $\mu_{f B}<1$

A figura 6.10 mostra os histogramas normalizados e as funções de massa do número de infectados em diferentes gerações com o contágio modelado por B = geométrica - 0 (0.4). Nela é possível observar a propagação da incerteza ao longo das gerações, e a adequação dos valores calculados ao modelo proposto.

O contágio de um indivíduo infectado não é limitado a um número restrito de indivíduos. Como, neste caso, $\mu_B < 1$ vemos que a média e a variância tendem a zero (equação 5-16 e figura 6.11a), indicando que, eventualmente, a epidemia irá se extinguir com 100% de probabilidade (figura 6.11b).



Histogramas e Funções de Massa de $I_n \operatorname{com} B = \operatorname{geométrica-0} (0.4)$

Figura 6.10: Comparação entre os histogramas normalizados da geométrica - 0 (0.4) construídos com 4000 realizações e as funções de massa calculadas analiticamente.



6.11(a): Estatísticas Amostrais.

6.11(b): Probabilidade de Extinção.

Figura 6.11: Contágio B = geométrica - 0 (0.4), com $\mu_B < 1$.

6.3.2 Contágio B= geométrica - 0 (0.5), com $\mu_{\rm B}=1$

A figura 6.12 mostra, novamente, os histogramas normalizados e as funções de massa do número de infectados em diferentes gerações com o contágio modelado por B = geométrica - 0 (0.5). Nela é possível observar a propagação da incerteza ao longo das gerações, e a adequação dos valores calculados ao modelo proposto.

A figura 6.13 mostra a média e variância do número de infectados ao longo das gerações e probabilidade de extinção com o contágio modelado por uma



Histogramas e Funções de Massa de $I_n \operatorname{com} B = \operatorname{geométrica-0} (0.5)$

Figura 6.12: Comparação entre os histogramas normalizados da geométrica - 0 (0.5) construídos com 4000 realizações e as funções de massa calculadas analiticamente.



6.13(a): Estatísticas Amostrais.

 $6.13(\mathrm{b})$: Probabilidade de Extinção.

Figura 6.13: Contágio B = geométrica - 0 (0.5), com $\mu_B = 1$.

geométrica - 0 (0.5). Como, neste caso, $\mu_B = 1$ vemos que a média permanece em torno de 1 enquanto a variância cresce linearmente conforme as gerações vão passando (equação 5-16 e figura 6.13a), indicando que, eventualmente, a epidemia irá se extinguir com 100% de probabilidade (figura 6.13b).

6.3.3 Contágio ${f B}=$ geométrica - 0 (0.6), com $\mu_{f B}>1$

A figura 6.14 mostra os histogramas normalizados e as funções de massa do número de infectados em diferentes gerações com o contágio modelado por

Capítulo 6. Influência do Modelo do Contágio no Comportamento da Doençã7

B = geométrica - 0 (0.6). Nela é possível observar, mais uma vez, a propagação da incerteza ao longo das gerações, e a adequação dos valores calculados ao modelo proposto.



Histogramas e Funções de Massa de $I_n \operatorname{com} B = \operatorname{geométrica-0} (0.6)$

Figura 6.14: Comparação entre os histogramas normalizados da geométrica - 0 (0.6) construídos com 4000 realizações e as funções de massa calculadas analiticamente.



Figura 6.15: Contágio $B = \text{geométrica} - 0 \ (0.6), \text{ com } \mu_B > 1.$

A figura 6.15 mostra a média e variância do número de infectados ao longo das gerações e probabilidade de extinção com o contágio modelado por uma geométrica - 0 (0.6). Assim como nos casos anteriores, o contágio não é limitado a um número de indivíduos No entanto, como neste caso, $\mu_B > 1$ vemos que a média e a variância crescem conforme as gerações vão passando, tendendo a infinito (equação 5-16 e figura 6.15a). Com base no equacionamento proposto em 5-18 temos, então, uma probabilidade de extinção $e \neq 1$. Para os parâmetros em questão, a probabilidade é e = 66.67% (figura 6.15b).

6.4 Modelo Probabilístico de B - Geométrica (p)

O terceiro modelo analisado foi a geométrica(p). Nesta análise, menos conservadora, um indivíduo infectado que sai de casa não tem um limite máximo de infectados, e sempre infecta no mínimo um novo indivíduo. O peculiar desta análise é: se todo indivíduo infectado contamina, pelo menos, uma nova pessoa essa epidemia não vai se extinguir. Esse modelo retrata uma sociedade com regras de distanciamento social baixas, com contato entre pessoas mais alto e menos restrito do que quando se modela o contágio com uma binomial ou geométrica-0.

Sua função geradora de probabilidade é dada por $G_B(s) = \frac{ps}{1-qs}$, e sua média é $\mu_B = \frac{1}{p}$. Para qualquer valor de $p \neq 0$ temos $\mu_B \geq 1$, confirmando que a epidemia nunca se extingue. Isto implica em probabilidade de extinção e = 0.

Histogramas com B = geométrica-1 (0.88)



Figura 6.16: Histogramas normalizados com o total de infectados nas gerações 5, 10, 15 e 20, construídos com 4000 realizações.

Abaixo estão os gráficos de média, variância e probabilidade de extinção analíticos e amostrais para o total de infectados em diferentes gerações quando temos o contágio modelado por B = geométrica - 1 (p).



6.17(a): Estatísticas Amostrais.

6.17(b): Probabilidade de Extinção.

Figura 6.17: Contágio B = geométrica - 1 (0.88), com $\mu_B > 1$.



Figura 6.18: Contágio B = geométrica - 1 (0.78), com $\mu_B > 1$.



Figura 6.19: Contágio B = geométrica - 1 (0.68), com $\mu_B > 1$.

7 Influência da Vacinação na Modelagem do Processo de Ramificação

Fundamentais para o combate a doenças na história da medicina, as vacinas estão também no epicentro de debates sobre tratamentos medicinais efetivos e leis compulsórias de imunização. Ao longo da história, elas ajudaram a reduzir expressivamente a incidência de pólio, sarampo e tétano, entre várias outras doenças. Hoje, são consideradas o tratamento com melhor custobenefício em saúde pública [17].

7.1 O Que São as Vacinas ?

As vacinas são preparações biológicas introduzidas no corpo humano a fim de protegê-lo e imunizá-lo de doenças. Elas ativam o sistema imunológico "ensinando"o corpo a reconhecer e combater vírus ou bactérias em futuras infecções. São compostas por agentes que se assemelham aos microrganismos causadores das doenças, e são muitas vezes feitas com formas enfraquecidas ou mortas do micróbio, das suas toxinas ou de uma das suas proteínas de superfície. Ao ser introduzida no corpo, a vacina estimula o sistema imunológico a produzir os anticorpos necessários para evitar o desenvolvimento da doença.

Os primeiros vestígios do uso de vacinas, com a introdução de versões atenuadas de vírus foram feitas em combate à varíola no século X, na China [34]. Em 1798 que o termo "vacina" surgiu oficialmente pela primeira vez. O médico e cientista Edward Jenner soube que trabalhadores da zona rural não se contaminavam com a varíola, pois haviam tido a varíola bovina. Ele então introduziu os dois vírus em uma criança e percebeu que ela, depois de reagir ao vírus, não tinha mais se infectado com a doença. A palavra vacina deriva justamente de Variolae vaccinae, nome científico dado à varíola bovina.

Em 1881, o cientista francês Louis Pasteur começou a desenvolver a segunda geração de vacinas, voltadas ao combate da cólera e do antraz [29]. A partir de então, as vacinas começaram a ser produzidas em massa e se tornaram um dos principais elementos para o combate a doenças, como a difteria, sarampo, caxumba e rubéola e pólio, ao redor do mundo.

7.2 Modelos Utilizados na Modelagem de uma Vacinação em Massa

Derivado do modelo proposto por Kermack [22], o modelo determinístico mais utilizado para a modelagem computacional de uma vacinação em massa é o SIRV. Modelo de compartimento composto por quatro grupos distintos. Susceptíveis, infectados, recuperados e vacinados. De modo geral, o grupo de vacinados diminui o tamanho do grupo de susceptíveis a doença. Devido às características do sistema, similares ao do modelo SIR explicado na seção 2.2.2, a curva característica de indivíduos infectados atinge um pico menor.

Derivado do modelo determinístico SIRV, diversos autores adaptaram a abordagem do modelo de compartimento para um modelo estocástico [9, 38, 37]. Modelos ainda mais complexos utilizaram o RAS, do inglês *realistic age-structured*, como Bolker [4] e Ferguson [16].

7.3 Construção de um Modelo Estocástico para uma População Parcialmente Vacinada

Dada uma população, modelaremos o contágio de um indivíduo da seguinte maneira. Baseado no modelo de ramificação da seção 5.1, um indivíduo infectado $(i_0 = 1)$, sai de sua casa e encontra m indivíduos. Supondo que em cada encontro o infectado pode ou não transmitir uma carga viral suficiente para contagiar o próximo, modelaremos o número de indivíduos que um indivíduo infectado encontrou e transmitiu a carga viral suficiente para contagiar com a variável aleatória C_j = binomial (m, p). Assim como o contágio da seção 5.1, todos os C_j são independentes e identicamente distribuídos (IID).

A cada contato com transmissão suficiente de carga viral, então, avaliaremos duas situações: primeiro, se esse indivíduo foi vacinado ou não; segundo, caso ele tenha sido vacinado, se a vacina foi efetiva ou não. Isto é, se ele está imunizado de fato, ou não. O primeiro evento, vacinação do indivíduo, será modelado por uma variável aleatória discreta V = Bernoulli (p_V) . O segundo evento, análise da efetividade da vacina, será modelado por outra variável aleatória discreta $E_l =$ Bernoulli (p_E) .

Em outras palavras, um indivíduo infectado entra em contato m indivíduos e transmite carga viral suficiente para contaminação de C indivíduos. Para cada um desses C novos indivíduos fazemos uma pergunta: "Esse indivíduo foi vacinado?". A probabilidade p_V determina a porcentagem de população vacinada. Se ele tiver sido vacinado nos perguntamos: "Será que a vacinação foi efetiva?". A eficácia da vacina é dada por p_E . Se foi, então a imunização funcionou, e este indivíduo está livre da doença. Se não foi, então ele está infectado. A figura 7.3 representa o esquema descrito.



Infectado.

Figura 7.1: Esquema de avaliação de um possível novo infectado.

Essa análise é feita para cada um dos C indivíduos que o primeiro infectado entra em contato e transmite carga viral suficiente para infectá-los.

A cada geração n, o total de indivíduos infectados I_n é dado pela quantidade de pessoas que entraram em contato com um indivíduo infectado e receberam carga viral suficiente para contagiar com a doença menos a quantidade de pessoas que tomaram a vacina e ela, de fato, foi eficaz.

Para modelar o novo cenário temos:

- $-I(n) = I_n$, número total de infectados em uma geração;
- C_j , número de indivíduos que entraram em contato com o j-ésimo indivíduo infectado da geração anterior e receberam carga viral suficiente para se contagiar com a doença, modelado por $C_j = \text{Binomial}(m, p)$, v.a. discreta;
- V_k , variável aleatória que determina se um indivíduo que teve contato com um infectado e recebeu carga viral suficiente para se contaminar está ou não está vacinado é modelada por V_K = Bernoulli (p_V) , v.a. discreta;
- $-E_l$ é a variável aleatória que determina se um indivíduo que teve contato com um infectado, recebeu carga viral suficiente para contaminação mas

recebeu a vacina se contamina ou não com a doença. E_l é modelado por E_l = Bernoulli (p_E) , v.a. discreta.

Todas os C_j são IID entre si, então $C_j \to C$. O mesmo vale para $V_k \in E_l$. Isto é $V_k \to V \in E_l \to E$.

As funções geradoras de C, V, E são:

$$G_V(x) = (1 - p_V) + p_V x.$$
 (7-1)

$$G_{E_l}(x) = (1 - p_E) + p_E x.$$
(7-2)

$$G_C(x) = ((1-p) + px)^m.$$
(7-3)

Influência da Vacinação no Comportamento da Epidemia

8.1 Influência da Vacinação

Para analisar a influência da vacinação no total de infectados em cada geração, na propagação da doença e nas probabilidades de extinção, foram feitas simulações de Monte Carlo e construídos modelos estatísticos para a propagação da doença considerando-se a v.a. C: binomial. Assim como no capítulo 6, 21 distribuições diferentes foram selecionadas. Para cada distribuição selecionada foram escolhidas 6 diferentes porcentagens de população vacinada (2%, 10%, 20%, 50%, 75% e 100%), e para cada uma dessas, foram escolhidas 4 valores de eficácia de uma vacina (50.4%, 62.1%, 70.4% e 95%). Para cada cenário descrito, 4.000 simulações do processo de ramificação da geração n = 0 a n = 20 foram computadas. No total, foram realizadas mais de 2 milhões realizações do processo estocástico de ramificação. Para fazer todas as simulações gastou-se aproximadamente 100 horas de processamento (tempo de CPU). O elevado custo computacional em tempo e volume de dados, caracteriza o problema como *biq data*. Um estudo de convergência foi desenvolvido para determinar o número de simulações [36]. Estatísticas amostrais (média, variância e probabilidade de extinção) e histogramas foram construídos com as respostas das realizações.

Antes de discutir a propagação da doença em uma população com vacinação, vale relembrar como é o comportamento do total de infectados em uma população não vacinada. Nas seções 6.2.1, 6.2.2 e 6.2.3 temos as 3 situações criíticas para um meio sem vacina. Para uma binomial com $\mu_B < 1$ temos a média e a variância tendendo a 0, indicando extinção da doença; para $\mu_B = 1$, média em torno de 1 e variância tendendo a 0, indicando também a extinção da doença; e para binomial com $\mu_B > 1$ temos média e variância tendendo a infinito, dando margem para uma probabilidade de extinção da doença diferente de 100%. Assim como mostra a figura 8.1.

Para um meio com vacina, onde foram vacinados 20% da população, com uma vacina de eficácia comprovada de 50.4% e com o contato C entre os indivíduos modelado por uma binomial(3, p), com os valores de p estipulados



8.1(a): Média amostral para os 3 valores crí- 8.1(b): Probabilidade de extinção para os 3 valores críticos de μ_C .

Figura 8.1: Gráficos construídos com 4000 realizações do comportamento sem vacina.

na seção 6.2, foi obtido o seguinte resultado (figura 8.2):



8.2(a): Média amostral para os 3 valores crí- 8.2(b): Probabilidade de extinção para os 3 valores críticos de μ_C .

Figura 8.2: Gráficos construídos com 4000 realizações do comportamento de um local com 20% da população vacinada com uma vacina com 50.4% de eficácia.

Comparando os dois conjunto de gráficos 8.1 e 8.2 vemos que: para $\mu_C < 1$ a epidemia se extingue com mais rapidez; para $\mu_C = 1$ a média deixa de flutuar em torno de 1 e passa a se comportar como uma função que tem a média $\mu_B < 1$. Além disso, temos para o mesmo intervalo de gerações a epidemia se extinguindo em menos tempo; para $\mu_C > 1$, neste caso, temos o decaimento da média amostral por geração em comparação com a situação não vacinada. Além disso, vamos que a probabilidade de extinção da doença passa de e = 23.6%para e = 35.7%.

E sempre interessante mostrar a propagação de incerteza no modelo proposto. Os histogramas da figura 8.3 mostram o cenário da curva na qual $\mu_C > 1$.



Figura 8.3: Histogramas normalizados com o total de infectados nas gerações 2, 4, 6 e 21, construídos com 4000 realizações.

8.2 Influência da Porcentagem da População Vacinada

Vimos que com 20% da população vacinada já temos uma mudaça significativa no comportamento da epidemia, em termos de pessoas infectadas e chances de ela se extinguir. No entanto, será que é possível mudarmos totalmente o resultado de uma epidemia apenas aumentando a quantidade de pessoas vacinadas?

Para tal, repetimos o mesmo processo da seção 8.1, avaliando diferentes valores de população vacinada. Dentre todos os valores calculados, estão mostrados a seguir populações que tiveram 2%, 50% e 75% de seus indivíduos vacinados com uma vacina de 50.4% de eficácia, com o contato C entre os indivíduos modelado por uma binomial(3, p).

Para apenas 2% da população vacinada é possível observar que as diferenças são pouco significativas. Para o cenário mais crítico ($\mu_C > 1$) ainda temos uma explosão no número de infectados e uma probabilidade de extinção baixa. A propagação de incerteza do modelo é observada nos histogramas das gerações 2, 4, 6 e 10, construídos com 4000 realizações na figura 8.4.

Já com 50% da população vacinada vemos um decaimento significativo no total de infectados. Contatos com média $\mu_C \leq 1$ se extinguem com 100% de probabilidade, além de se extinguirem com menos gerações. Para contatos com $\mu_C > 1$ vemos o decaimento mais significativo até então. A média de infectados diminui de 4000 infectados para apenas 10 e a probabilidade de extinção aumenta em aproximadamente 200%, indo de 23.6% para 67.4%. A propagação de incerteza do modelo é observada nos histogramas das gerações 2, 4, 6 e 21 na figura 8.6.



Figura 8.4: Histogramas normalizados com o total de infectados nas gerações 2, 4, 6 e 10, construídos com 4000 realizações.



8.5(a): Média amostral para os 3 valores críti- 8.5(b): Probabilidade de extinção para os 3 valores críticos de μ_C .

Figura 8.5: Gráficos construídos com 4000 realizações do comportamento de um local com 2% da população vacinada com uma vacina com 50.4% de eficácia.

Para 75% da população vacinada, então, temos a total conversão do caso crítico. Uma situação onde teríamos uma contaminação em massa, temos agora a extinção total do vírus. Apesar de no gráfico 8.9(b) ainda não termos a estabilização da probabilidade de extinção, é possível concluir que a epidemia se extinguirá, isto é, e = 1 devido ao comportamento característico de decaimento da curva de média de infectado (figura 8.9(a)). A propagação de incerteza do modelo é observada nos histogramas das gerações 2, 4, 6 e 21, construídos com 4000 realizações na figura 8.8.



Figura 8.6: Histogramas normalizados com o total de infectados nas gerações 2, 4, 6 e 21, construídos com 4000 realizações.



8.7(a): Média amostral para os 3 valores críti- 8.7(b): Probabilidade de extinção para os 3 valores críticos de μ_C .

Figura 8.7: Gráficos construídos com 4000 realizações do comportamento de um local com 50% da população vacinada com uma vacina com 50.4% de eficácia.

8.3 Influência da Eficácia da Vacina

Outro fator de análise é a eficácia da vacina aplicada. Vimos que com uma vacina com 50.4% de eficácia temos uma mudança significativa no comportamento da epidemia, principalmente com a vacinação de pelo menos 75% da população. No entanto, será que também é possível mudarmos totalmente o resultado de uma epidemia vacinando menos gente, mas com uma vacina mais eficaz?

Para tal, repetimos o mesmo processo da seção 8.1, avaliando diferentes valores de eficácia de vacina. Dentre todos os valores de eficácia simulados, está a seguir apenas os resultados para 50.4%, 62.1% e 70.4% de eficácia de



Figura 8.8: Histogramas normalizados com o total de infectados nas gerações 2, 4, 6 e 21, construídos com 4000 realizações.



8.9(a): Média amostral para os 3 valores críti- 8.9(b): Probabilidade de extinção para os 3 valores críticos de μ_C .

Figura 8.9: Gráficos construídos com 4000 realizações do comportamento de um local com 75% da população vacinada com uma vacina com 50.4% de eficácia.

uma vacina. Além disso, dentre as porcentagens de população vacinadas, será mostrado apenas os resultados das simulações com porcentagens de 50% e 75% de população vacinadas. Em todos os casos, temos o contato C entre os indivíduos modelado por uma binomial(3, p).

8.3.1 Vacinando 50% da População

A partir da figura 8.11, $p_E = 0.5$, observamos que houve um decaimento na média de infectados. No entanto, para o contágio modelado por uma binomial com média $\mu_C > 1$ ainda podemos observar que a epidemia ainda não se extinguirá.

A figura 8.10 mostra a propagação de incerteza no decorrer das gerações.



Figura 8.10: Histogramas normalizados com o total de infectados nas gerações 2, 4, 6 e 21, construídos com 4000 realizações.



8.11(a): Média amostral para os 3 valores críti- 8.11(b): Probabilidade de extinção para os 3 valores críticos de μ_C .

Figura 8.11: Gráficos construídos com 4000 realizações do comportamento de um local com 50% da população vacinada com uma vacina com 50.4% de eficácia.

O histograma da geração 21 confirma que, apesar do decaimento, ainda teremos indivíduos infectados com a doença.

A figura 8.13 mostra o cenário de 50% da população vacinada com uma vacina de eficácia de 62%. Observamos que também houve um decaimento na média de infectados. Para o contágio modelado por uma binomial com média $\mu_C > 1$ ainda podemos observar que a epidemia ainda não se extinguirá. Comparando as figuras 8.11(b) e 8.13(b) nota-se que ao menos com uma vacina mais eficaz aumentamos a probabilidade de extinção dela. Ainda assim, nenhum dos dois casos garante a extinção da doença.

É possível observar na figura 8.12, principalmente no histograma da geração 21, que o suporte da v.a. é menor do que quando comparado com o da figura 8.10. Apesar de ainda não termos a extinção, já temos uma melhora



Figura 8.12: Histogramas normalizados com o total de infectados nas gerações 2, 4, 6 e 21, construídos com 4000 realizações.



8.13(a): Média amostral para os 3 valores críti
- 8.13(b): Probabilidade de extinção para os 3 valores críticos de
 μ_C .

Figura 8.13: Gráficos construídos com 4000 realizações do comportamento de um local com 50% da população vacinada com uma vacina com 62.1% de eficácia.

no quadro geral. O total de infectados no final das gerações é menor com uma vacina mais eficaz.

A figura 8.15 mostra o cenário de 50% da população vacinada com uma vacina de eficácia de 70.4%. Observamos que também houve um decaimento na média de infectados. Neste caso, observamos que para todos os valores críticos temos a conversão de todas as modelagens para um cenário onde as epidemias irão se extinguir. Apesar das 3 curvas decaírem, a curva do contato modelado por uma binomial de média maior que 1 decai lentamente. Pelo gráfico da figura 8.15(b) e pelos histogramas da figura 8.14 vemos que com 20 gerações a epidemia ainda não se extingue, apesar de tender a 100% de probabilidade de extinção.



Figura 8.14: Histogramas normalizados com o total de infectados nas gerações 2, 4, 6 e 21, construídos com 4000 realizações.



8.15(a): Média amostral para os 3 valores críti- 8.15(b): Probabilidade de extinção para os 3 valores críticos de μ_C .

Figura 8.15: Gráficos construídos com 4000 realizações do comportamento de um local com 50% da população vacinada com uma vacina com 70.4% de eficácia.

8.3.2 Vacinando 75% da População

O último grupo de cenários analisados retrata uma sociedade com 75% da população vacinada. Para cada um dos 3 cenários cíticos ($\mu_C < 1$, $\mu_C = 1$ e $\mu_C > 1$), serão mostradas apenas as simulações com 3 valores diferentes de eficácia da vacina (50.4%, 62.1% e 70.4%). Em todos os casos simulados, figuras 8.17, 8.19 e 8.21, obtivemos o comportamento desejado: todos os cenários tiveram suas médias decaindo com o passar das gerações e as probabilidades de extinção tendendo a e = 1.

Os histogramas das figuras 8.16, 8.18 e 8.20 confirmam os resultados achados: para uma população inicial de $i_0 = 1$, com 75% da população vacinada
teremos o fim da epidemia.

75% da população vacinada, 50.4% de eficácia, C = binomial (3, 0.5)



Figura 8.16: Histogramas normalizados com o total de infectados nas gerações 2, 4, 6 e 21, construídos com 4000 realizações.



8.17(a): Média amostral para os 3 valores críti
- 8.17(b): Probabilidade de extinção para os 3 valores críticos de
 $\mu_C.$

Figura 8.17: Gráficos construídos com 4000 realizações do comportamento de um local com 75% da população vacinada com uma vacina com 50.4% de eficácia.



75% da população vacinada, 62.1% de eficácia, C = binomial (3, 0.5)

Figura 8.18: Histogramas normalizados com o total de infectados nas gerações 2, 4, 6 e 21, construídos com 4000 realizações.



8.19(a): Média amostral para os 3 valores críti
- 8.19(b): Probabilidade de extinção para os 3 valores críticos de
 $\mu_C.$

Figura 8.19: Gráficos construídos com 4000 realizações do comportamento de um local com 75% da população vacinada com uma vacina com 62.1% de eficácia.



75% da população vacinada, 70.4% de eficácia, C= binomial (3,0.5)

Figura 8.20: Histogramas normalizados com o total de infectados nas gerações 2, 4, 6 e 21, construídos com 4000 realizações.



8.21(a): Média amostral para os 3 valores críti
- 8.21(b): Probabilidade de extinção para os 3 valores críticos de
 $\mu_C.$

Figura 8.21: Gráficos construídos com 4000 realizações do comportamento de um local com 75% da população vacinada com uma vacina com 70.4% de eficácia.

9 Conclusão

Este trabalho analisou com uma abordagem estocástica a propagação de uma doença epidemiológica em uma população. Na primeira parte do trabalho foi feita uma análise determinística do processo de propagação de uma doença. Para tal, foram utilizados os modelos de compartimentos propostos por McKendric e Kermac [22]. O primeiro modelo de compartimentos estudado foi o modelo SI, composto apenas por dois compartimentos (susceptíveis e infectados). Esse modelo tem como característica a contaminação de todos os indivíduos da sociedade, independente do valor da taxa de contágio β . O segundo modelo de compartimentos foi o SIR. Mais complexo que o SI, este modelo apresenta um terceiro grupo: recuperados, compostos pela porcentagem de indivíduos que não podem mais se contaminar. Os parâmetros deste modelo são β , taxa de contágio; α , taxa de recuperação; e r_o , número básico de reprodução, razão entre $\beta e \alpha$. Diferentes valores de r_o resultaram em curvas com comportamentos distintos, principalmente quando avaliados em relação ao total de indivíduos infectados.

Como os métodos de determinação do r_o podem resultar em diferentes valores para o total de indivíduos contaminados e, consequentemente, diferentes propagações da doença no meio da sociedade, surgiu a oportunidade de olhar o contágio de uma entre endivíduos e a propagação sob uma ótica estocástica.

Na segunda parte do trabalho tivemos a introdução de alguns conceitos básicos de probabilidade, junto com a nova modelagem proposta de análise da propagação de uma doença. O número de indivíduos infectados por cada membro infectado da população foi modelado por uma variável aleatória discreta e o número de indivíduos infectados ao longo do tempo foi modelado por um processo estocástico de ramificação. Para caracterizar a propagação das incertezas na disseminação da doença, foram construídos histogramas do número de indivíduos infectados. Estatísticas amostrais também foram calculadas. Foi feita uma análise da influência do modelo probabilístico da variável que modela o contágio entre os indivíduos no processo de ramificação. Foram calculados também, as probabilidades de extinção para as três famílias diferentes de v.a. Os resultados desta parte do trabalho mostraram que os parâmetros do modelo probabilístico da v.a. que modela o contágio, B, influenciam fortemente o comportamento da propagação de uma doença. B sendo uma binomial(m, p), o número de indivíduos infectados por cada membro já infectado é limitado a m indivíduos. Quando B foi modelado por uma geométrica-1 (p) ou uma geométrica-0 (p), o número de novos indivíduos contagiados não teve limite máximo. Como foi visto, para a B = binomial(m, p) e B = geométrica-0(p), para valores de $\mu_B \leq 1$ a epidemia se extinguiu com probabilidade e = 1. Para v.a. com valores de $\mu_B > 1$, a probabilidade de extinção assumiu valores entre $e \in [0, 1]$. Para a distribuição geométrica-1 vimos que, independente do valor de p, tivemos uma distribuição com média $\mu_B > 1$. Além disso, como ela retrata uma sociedade onde a cada geração temos pelo menos um novo contágio, tivemos que a probabilidade de extinção é e = 0.

Neste trabalho também foi feito um estudo sobre as soluções analíticas deste problema. A partir das funções geradoras apresentadas calculou-se as médias, variâncias e probabilidades de extinção analíticas. Além dessas estatísticas, foi possível obter a função de massa do total de infectados em cada geração. Comparando as distribuições analíticas e amostrais foi possível validar o processo de geração de amostras do processo de ramificação através do método de Monte Carlo e complementar a analise de convergência feita para determinar o número de amostras necessárias para estimar estatísticas amostrais dentro de uma margem de erro prescrita.

A terceira e última parte trouxe a propagação de uma doença em uma população parcialmente vacinada. Dado que uma parcela da população esta vacinada, modelamos o contato capaz de transmitir carga viral por uma v.a. discreta, C = binomial(m, p); a determinação se um indivíduo que teve contato com um infectado esta ou não vacinado foi modelada por outra v.a discreta, $V = \text{bernoulli}(p_V)$; a eficácia da vacina foi modelada por uma v.a. discreta, $E = \text{bernoulli}(p_E)$. O número total de indivíduos infectados ao longo do tempo foi modelado por um processo de ramificação estocástico (I_n) . Para caracterizar a propagação das incertezas na disseminação da doença, foram construídos histogramas do número de indivíduos infectados. Estatísticas amostrais também foram calculadas. Foi feita uma análise da influência da eficácia da vacina no processo de ramificação, bem como da influência da porcentagem da população vacinada. Foram calculadas também, as probabilidades de extinção para os diferentes casos analisados.

Os resultados obtidos na seções 8.2 e 8.3 mostram que é possível controlar e mitigar uma epidemia com uma vacinação adequada. Apesar da eficácia da vacina aplicada interferir no total de infectados ao longo do tempo, ela pode ou não garantir a extinção da doença.

Este trabalho gerou dois artigos de congresso e um artigo de revista. O primeiro, intitulado Análise Estocástica da Propagação de Doenças Epidemiológicas [5], foi publicado no X Encontro Regional de Matemática Aplicada e Computacional do Rio Grande do Sul (X ERMAC-RS), realizado entre os dias 1 e 3 de dezembro de 2020. O segundo, How the spread of an infectious disease is affected by the contagion's probabilistic model [6], publicado no XIV Encontro Acadêmico de Modelagem Computacional, realizado entre os dias 8 e 11 de fevereiro de 2021. O terceiro artigo publicado foi decorrente da publicação do artigo do X-ERMAC. Os PDFs desses artigos encontram-se no anexo, ao final desta dissertação

Referências bibliográficas

- ALLEN, L.. An Introduction to Stochastic Processes with Applications to Biology, v. 2. Taylor & Francis, USA, 2010.
- [2] BERNOULLI, D.. Réflexions Sur Les Avantages de L'inoculation. Mercure de France, 1:173–190, 1760.
- [3] BERNOULLI, D.. Essai d'une Nouvelle Analyse de la Mortalité Causeé par la Petite Vérole. Mem. Math. Phys. Acad. Sci. Paris, 1:1–45, 1766.
- [4] BOLKER, B. M. GREFNELLL, B. T.. Impact of Vaccination on the Spatial Correlation and Persistence of Measles Dynamics. Proc. Natl. Acad. Sci. USA, 93(22):12648–12653, 1996.
- [5] BORGES, B.; LIMA, R. ; SAMPAIO, R.. Análise Estocástica de Propagação de Doenças Epidemiológicas. In: Anais do X encontro regional de matemática aplicada e computacional do rio grande do sul – ermac-rs, p. 1–8, Brasil, 2020.
- [6] BORGES, B.; LIMA, R. ; SAMPAIO, R.. How the Spread of an Infectious Disease is Affected by the Contagion's Probabilistic Model. In: Anais do xiv encontro acadêmico de modelagem computacional., p. 6–16, Brasil, 2021.
- [7] BRAUER, F.; CASTILLO-CHAVEZ, C.; FENG, Z. Mathematical Models in Epidemiology, v. 69. Springer, USA, 2019.
- [8] BUCK, C. LOPIS, E.; NAJERA, E.. The Challenge of Epidemiology: Issues and Selected Readings. Pan American Health Organization, England.
- CHEN, C. KANG, Y.. The Asymptotic Behavior of a Stochastic Vaccination Model with Backward Bifurcation. Applied Mathematical Modelling, 40(2016):6051 – 6068, 2016.
- [10] DICKERS, R.. Principles of epidemiology: an introduction to applied epidemiology and biostatistics, v. 2. Centers for Disease

Control and Prevention (U.S.), Public Health Practice Program Office., Atlanta, GA, 2006.

- [11] DIEKMANN, O.; HEESTERBEEK, H.; BRITTON, T. Mathematical Tools for Understanding Infectious Disease Dynamics, v. 1. Princeton University Press, Oxford, GBR, 2013.
- [12] DIETZ, K. HEESTERBEEK, J.. Daniel Bernoulli's epidemiological model revisited. Math. Biosc., 180(2):1–221, 2002.
- [13] DOHERTY, P. C.. Pandemics, What Everyone Need to Know., v. 1. Oxford University Press, Oxford, GBR, 2013.
- [14] FARR, W.. Progress of Epidemics. Second Report of the Registrar General of England and Wales, 2(1):91–98, 1840.
- [16] FINE, P., Herd Immunity: History, Theory, Practice. Epidemiologic Reviews, 15(2):265–302, 1993.
- [16] FERGUSON, N. ANDERSON, R.; GARNETT, G. Mass Vaccination to Control Chickenpox: The Influence of Zoster. Proc. Natl. Acad. Sci. USA, 93(14):7231-7235, 1996.
- [17] FIOCRUZ. Vacinas: As Origens, a Importância e os Novos Debates Sobre Seu Uso. Instituto de Tecnologia em Imunobiológicos (Bio-Manguinhos): Fundação Oswaldo Cruz (Fiocruz), 2016. Acesso em: Julho de 2016.
- [18] GRAUNT, J.. Natural and Political Observations Made upon the Bills of Mortality, v. 1. Fellow of the Royal Society., London, GBR, 1676.
- [19] GRIMMETT, G.; WELSH, D.. Probability: An Introduction, v. 2. Oxford University Press, Oxford, GBR, 2014.
- [20] HAMER, W.. Epidemic Disease in England: the Evidence of Variability and of Persistence of Type. The Lancet, 167(4305):733-738, 1906.
- [21] HACCOU, P.; HACCOU, P.; JAGERS, P. ; VATUTIN, V.. Branching Processes: Variation, Growth, and Extinction of Populations, v. 1. Cambridge University Press, Cambridge, GBR, 2005.
- [22] KERMACK, W. MCKENDRIC, A. A Contribution to the Mathematical Theory of Epidemics. Proceedings to the Royal Society of London, Series A, 115(772):700-721, 1927.

- [23] LIMA, R.; SAMPAIO, R.. What is Uncertainty Quantification? Journal of the Brazilian Society of Mechanical Sciences and Engineering, 3(40):155, 2018.
- [24] LIMA, R.; SAMPAIO, R.: How to Deal with Uncertainty Quantification and Propagation. Mecânica Computacional, 16(XXXVI):723–739, 2018.
- [25] MARTCHEVA, M.. An Introduction to Mathematical Epidemiology, v. 61. Springer US, New York, USA, 2015.
- [26] NELSON, K. E. WILLIANS, C.. Infectious Desease Epidemiology: Theory and Practice, v. 3. Jones e Barlett Learning, Burlington, MA, 2014.
- [27] OLOFSSON, P. ANDERSSON, M.. Probability, Statistics, and Stochastic Processes., v. 2. John Wiley and Sons INC., New York, USA, 2011.
- [28] OLSEN, S.; MACKINON, L.; GOULDING, J.; BEAN, N. ; SLUTSKER, L.. Surveillance for Foodborne Disease Outbreaks in United States, 1993 to 1997. Surveillance Summaries, 49(SS-1), 2000.
- [29] PASTEUR, L.. Translation of an Address on the Germ Theory. The Lancet, 118(3024), 1881.
- [30] POLYA, G.. Mathematics and Plausible Reasoning, v. 1. Princeton University Press, London, GBR, 1954.
- [31] QUN, L.. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus Infected Pneumonia. The New England Journal of Medicine, 382(13):1199–1207, 2020.
- [32] SAMPAIO, R.; LIMA, R.. Modelagem Estocástica e Geração de Amostras de Variáveis e Vetores Aleatórios. In: Notas em matemática aplicada, v. 70, p. 144, Brasil, 2012.
- [33] SAVI, M.; SAVI, P. ; BORGES, B. A Mathematical Description of the Dynamics of Coronavirus Disease 2019 (COVID-19): A Case Study of Brazil. Computational and Mathematical Methods in Medicine, 2020:8, 2020.
- [34] SILVERSTEIN, A. M.. A History of Immunology, v. 2. Elsevier Academic Press., Oxford, GBR, 2009.

- [35] SNOW, J.. The Mode of Communication of Cholera., v. 1. Churchill, London, 1855.
- [36] SOUZA, DE CURSI, J.; SAMPAIO, R.. Modelagem Estocástica e Quantificação de Incertezas. Notas em Matemática Aplicada, 66:344, 2012.
- [37] WITBOOI, P. J.; MULLER, G. E.; VAN SCHALKWYK, G. J.. Vaccination Control in a Stochastic SVIR Epidemic Model. Computational and Mathematical Methods in Medicine, 2015:9, 2015.
- [38] YUAN, E. C.; ALDERSON, D. L.; STROMBERG, S. ; CARLSON, J. M.. Optimal Vaccination in a Stochastic Epidemic Model of Two Non-Interacting Populations. PLoS ONE, 10(2):25, 2015.

A Artigos Publicados

A seguir estão os PDFs dos artigos baseados nesta dissertação que já foram publicados.

O primeiro, intitulado Análise Estocástica da Propagação de Doenças Epidemiológicas, foi publicado no X Encontro Regional de Matemática Aplicada e Computacional do Rio Grande do Sul (X ERMAC-RS), realizado entre os dias 1 e 3 de dezembro de 2020.

O segundo, *How the spread of an infectious disease is affected by the contagion's probabilistic model*, publicado no XIV Encontro Acadêmico de Modelagem Computacional, realizado entre os dias 8 e 11 de fevereiro de 2021.

O terceiro, Análise Estocástica da Propagação de Doenças Epidemiológicas, foi publicado na Revista Mundi, Engenharia Tecnologia e Gestão, baseado no artigo de mesmo nome publicado no ano de 2020.

ANÁLISE ESTOCÁSTICA DE PROPAGAÇÃO DE DOENÇAS EPIDEMIOLÓGICAS

Beatriz Borges¹, Roberta Lima², Rubens Sampaio³

¹Pontifícia Universidade Católica do Rio de Janeiro, beatrizrbborges@gmail.com ²Pontifícia Universidade Católica do Rio de Janeiro, robertalima@puc-rio.br ³Pontifícia Universidade Católica do Rio de Janeiro, rsampaio@puc-rio.br

RESUMO: Este trabalho analisa com uma abordagem estocástica a propagação de uma doença epidemiológica em uma população. O foco é caracterizar a evolução do número de infectados ao longo do tempo, de forma a caracterizar a propagação de incertezas no número de infectados. Para fazer essa análise, o número de indivíduos contaminados por cada membro já infectado da população é modelado como uma variável aleatória binomial(m,p) e a propagação da doença é modelada como um processo estocástico de ramificação. Um modelo estatístico é construído para esse processo de ramificação. Histogramas e estatísticas amostrais do número de infectados ao longo do tempo, como por exemplo, média e variância, são calculadas através de simulações de Monte Carlo para o processo de propagação da doença. Os objetivos do trabalho são comparar as estatísticas amostrais calculadas com previsões da literatura, e determinar a influência do modelo probabilístico da variável aleatória de contágio no processo de ramificação. Através dessas análises deseja-se determinar qual é a probabilidade de extinção da doença.

Palavras-chave: Modelagem Estocástica; Propagação de Incertezas; Epidemiologia; Doenças Contagiosas.

1. INTRODUÇÃO

Propagação de doenças epidemiológicas é um tema que vem atraindo interesse de pesquisadores de diferentes áreas. Saber como doenças se propagam, pode facilitar o planejamento e implementação de medidas de controle e contenção. Além disso, prever o número de infectados ao longo do tempo permite que governos coordenem ações e melhor atendam a população, por exemplo, adequando o número de leitos hospitalares ao número de infectados.

Na literatura existem diversos modelos matemáticos para modelar a propagação de doenças, como por exemplo, o modelo de compartimentos de susceptíveis e infectados (SI) e o modelo de compartimentos de susceptíveis, infectados e recuperados (SIR). Em cada um desses modelos, a população é dividida em grupos (chamados de compartimentos) e um problema de valor inicial caracteriza a evolução do número de indivíduos em cada um desses grupos MARCHEVA (2015); BRAUER *et al.* (2019); LI (2018).

Apesar de amplamente utilizados em previsões, esses modelos abordam o processo de propagação de doenças de um ponto de vista determinístico, quando na verdade o problema é estocástico. A incerteza no contágio faz com que o número de infectados ao longo do tempo seja aleatório. Para fazermos melhores previsões da propagação, devemos levar em consideração essa estocasticidade.

Este trabalho analisa a propagação de uma doença epidemiológica entre indivíduos de uma população sob uma ótica estocástica. O número de indivíduos contaminados por cada membro já infectado da população é modelado por uma variável aleatória e o número de indivíduos infectados ao longo do tempo é modelado por um processo estocástico de ramificação. O foco do trabalho é caracterizar a propagação de incertezas no número de infectados, ou seja caracterizar como as funções de massa de seções do processo variam ao longo do tempo LIMA; SAMPAIO (2018); LIMA; SAMPAIO (2018). Para fazer essa análise, um modelo estatístico é construído para esse processo de ramificação. Histogramas e estatísticas amostrais do número de infectados ao longo do tempo, como por exemplo, média e variância, são calculadas através de simulações de Monte Carlo para o processo de propagação da doença. As simulações são feitas através do software Matlab. Os objetivos do trabalho são comparar as estatísticas amostrais calculadas com previsões da literatura, e determinar a influência do modelo probabilístico da variável aleatória de contágio no processo de ramificação. Através dessas análises deseja-se determinar qual é a probabilidade de extinção da doença.

2. CONSTRUÇÃO DE MODELOS PROBABILÍSTICOS PARA O CONTÁGIO E PROPAGAÇÃO DA DOENÇA

Esta seção trata da modelagem probabilística do contágio entre indivíduos da população e da propagação da doença. O número de indivíduos contaminados por cada membro já infectado j da população é modelado como uma variável aleatória (v.a.) discreta binomial(m,p) denominada C_j . Supõe-se que as v.a. C_j sejam independentes e identicamente distribuídas (IID). Com isso, denominamos C a variável aleatória que modela o contágio. A distribuição binomial pode ser interpretada, de acordo com conceitos de epidemiologia, como um experimento no qual um indivíduo encontra com um número m de pessoas, com uma probabilidade p de infectar cada uma delas. Depois dos encontros, avaliamos se cada uma das m pessoas foi, ou não, infectada e contamos o número de novos contagiados.

O número de infectados ao longo do tempo é modelado por um processo es-

tocástico I de parâmetro discreto chamado de processo de ramificação. Dado um conjunto de parâmetros discretos $N = \{0,1,2,3,...\}$, representando gerações ao longo do tempo, para cada $n \in N$, $I(n) = I_n$ é uma variável aleatória discreta. I_n representa o número de infectados na n-ésima geração e é dado pela soma do número de pessoas que foram contagiadas por cada um dos infectados na geração anterior n-1, ou seja:

$$I_n = \sum_{k=0}^{I_{n-1}} C_k , \qquad n > 0 .$$
 (1)

A Figura 1 ilustra as cinco primeiras gerações de uma realização do processo de ramificação I. Nesta realização, o processo de contaminação começa com um indivíduo infectado, $i_0 = 1$. Na geração n = 1, aparecem 3 indivíduos contaminados pelo indivíduo infectado inicialmente, $i_1 = 3$. Esse número de infectados é obtido fazendo-se uma realização da variável aleatória C. Na próxima geração, n = 2, o número total de infectados é a soma do número de pessoas que foram contagiadas por cada um dos três infectados na geração anterior. Para obter esse número, geramos três realizações independentes de C e somamos os valores sorteados. No exemplo ilustrado, o primeiro infectado contaminou apenas um indivíduo e os outros dois não fizeram nenhuma contaminação. Dessa forma $i_2 = 1$. Na próxima geração, n = 3, o único indivíduo contaminado na geração anterior contaminou quatro novos indivíduos, fazendo com que $i_3 = 4$. Na geração n = 4, o número total de infectados, que é a soma do número de pessoas que foram contagiadas por cada um dos quatro infectados em n = 3, é obtido gerando-se quatro realizações independentes de C e somando os valores sorteados. No exemplo ilustrado, o primeiro infectado contaminou dois indivíduos, o segundo infectado não fiz nenhuma contaminação e o terceiro e quarto fizeram duas contaminações cada. Totalizando $i_4 = 7$. Se nenhum desses sete indivíduos contaminados fizer novas contaminações, teremos $i_n = 0, \forall n \ge 5$, e nesse caso a doença estará extinta. A probabilidade de extinção de uma doença é uma das variáveis de interesse desse trabalho.

A função geradora de probabilidade de C é:

$$G_C(x) = \sum_{k=0}^{\infty} p_k x^k , \qquad (2)$$

onde p_k é a probabilidade de C assumir o valor k. Sendo C uma a variável aleatória binomial(m,p), a função geradora de probabilidade de C pode ser reescrita como:

$$G_C(x) = ((1-p) + px)^m$$
. (3)



Figura 1. Uma realização do processo de ramificação que modela o número de infectados ao longo de geraç oes. Considera-se $i_0 = 1$. Fonte: GRIMMET; WELSH (2014)

De acordo com o teorema 4.36 de GRIMMET; WELSH (2014), a função geradora de probabilidade de I_n é dada por:

$$G_{I_n}(x) = G_C(G_{I_{n-1}}(x)).$$
(4)

3. ESTATÍSTICAS DA PROPAGAÇÃO DA DOENÇA

Esta seção mostra algumas estatísticas do processo estocástico que modela a propagação da doença HACCOU *et al.* (2017); ALLEN (2010). De acordo com o teorema 9.8 de GRIMMET; WELSH (2014), a média de infectados ao longo das gerações é:

$$E(I_n) = \mu^n , \qquad (5)$$

onde $\mu = mp$ é a média da variável aleatória C, binomial(m,p). A longo prazo:

$$\lim_{n \to \infty} E(I_n) \to \begin{cases} 0 \text{ se } \mu < 1 ,\\ 1 \text{ se } \mu = 1 ,\\ \infty \text{ se } \mu > 1 . \end{cases}$$
(6)

Fazendo uma adaptação do teorema 9.8 de GRIMMET; WELSH (2014), podese escrever:

$$var(I_n) = \begin{cases} n\sigma^2 & \text{se } \mu = 1, \\ \sigma^2 \mu^{n-1} \left(\frac{\mu^n - 1}{\mu - 1}\right) & \text{se } \mu \neq 1, \end{cases}$$
(7)

onde $\sigma^2 = mp(1-p)$ é a variância de C. Analisando a variância a longo prazo:

$$\lim_{n \to \infty} var(I_n) \to \begin{cases} 0 \text{ se } \mu < 1 ,\\ \infty \text{ se } \mu \ge 1 . \end{cases}$$
(8)

A partir da análise mostrada em (6) e (8), temos que se $\mu \leq 1$ a epidemia, eventualmente, chegará ao fim. Para $\mu > 1$, por sua vez, a epidemia pode, ou não, chegar ao fim. Para determinar essa probabilidade de extinção definimos dois eventos distintos. Primeiro definimos um evento A: extinção da doença na *n*-ésima geração, i.e., $I_n = 0$. Utilizando a Equação (4) e avaliando-a em x = 0 temos:

$$G_{I_n}(x) = G_C(G_{I_{n-1}}(x))$$

$$G_{I_n}(0) = G_C(G_{I_{n-1}}(0)) = e_n ,$$
(9)

assim, chamamos de e_n a probabilidade de A. Depois, definimos o evento B: extinção da doença. De acordo com o teorema 1.54 de GRIMMET; WELSH (2014), a probabilidade deste evento ocorrer é dada pelo $\lim_{n\to\infty} e_n$ que será representado por e. Assim:

$$e_n = G_{I_n}(0) = G_C(G_{I_{n-1}}(0)) \quad \rightarrow \quad e_n = G_C(e_{n-1})$$
$$\lim_{n \to \infty} e_n = e \quad \rightarrow \quad e = G_C(e) , \tag{10}$$

onde a menor raiz da equação $e = G_C(e)$ no intervalo [0,1] é a probabilidade da epidemia se extinguir.

4. ESTATÍSTICAS AMOSTRAIS, PROBABILIDADE DE EXTINÇÃO E HISTOGRAMAS NORMALIZADOS

Para construir um modelo estatístico por Monte Carlo para a propagação da doença de forma poder compará-lo com o modelo encontrado na literatura, mostrado na seção 3, foram feitas 2000 realizações de I da geração n = 0 até n = 20através do software *Matlab*. Um estudo de convergência foi desenvolvido para determinar o número de realizações utilizadas SOUZA DE CURSI; SAMPAIO (2012). Estatísticas amostrais (média, variância e probabilidade de extinção) e histogramas foram construídos com as realizações.

Com o objetivo de analisar o comportamento estocástico da epidemia ao longo das gerações, foram construídos os histogramas normalizados do número de infectados para diferentes gerações do processo de ramificação com o contagio modelado pela binomial(3,0.5). Os resultados estão mostrados na Figura 2. Nessa figura, é possível observar a propagação de incertezas SAMPAIO; LIMA (2012) na propagação da doença, i.e., a evolução da função de massa do número de infectados ao longo das gerações.



Figura 2. Histogramas normalizados do número de infectados nas gerações 5, 10, 15 e 20, construídos com 2000 realizações da binomial(3,0.5). Fonte: do autor

A Figura 3 mostra a média e variância amostrais do processo I ao longo das gerações para diferentes valores de p. Considera-se $p = 0.16, 0.33 \in 0.50$. Observamos que os valores amostrais estão de acordo com os resultados analíticos dados pelas Equações (6) e (8).



Figura 3. Média e variância, analíticas e amostrais, para os valores p: 0.16, 0.33 e 0.50. Fonte: do autor

Para p = 0.16, com $\mu = 0.5 < 1$, podemos observar pela Figura 3(a) que tanto a média quanto a variância do processo estocástico tendem a zero, confirmando que para $\mu < 1$ temos a extinção da epidemia. Para p = 0.33, com $\mu = 1.0$, podemos observar pela Figura 3(b) que a média do processo fica em torno de 1, enquanto a variância cresce linearmente ao longo das gerações. Esse resultado está de acordo com a Equação (7). Para p = 0.50, com $\mu = 1.5 > 1$, podemos observar pela Figura 3(c) que tanto a média quanto a variância do processo crescem ao longo das gerações. Esse comportamento indica que $\lim_{n\to\infty} E(I_n) = \infty$ e $\lim_{n\to\infty} var(I_n) = \infty$, de forma que não há possibilidade da epidemia se extinguir.

De acordo com a Equação (10), a probabilidade de extinção da doença, e, é dada por $G_C(e) = ((1 - p) + pe)^m = e$. A Figura 4 mostra as probabilidades de extinção amostrais ao longo das gerações para os diferentes valores de p fixados acima. Observe que, com probabilidade e = 1, a epidemia tende a se extinguir para valores de $\mu \leq 1$, e que, para valores de $\mu > 1$, ela se extingue com valores de $e \in [0,1]$.



Figura 4. Gráfico da probabilidade de extinção da doença, analítica e amostral para os diferentes valores p: 0.16, 0.33 e 0.50. Fonte: do autor

5. CONCLUSÕES

Este trabalho utilizou uma abordagem estocástica para modelar o problema de propagação de uma doença epidemiológica em uma população. O número de indivíduos contaminados por cada membro infectado da população foi modelado por uma variável aleatória (v.a.) discreta binomial(m,p) e o número de infectados ao longo do tempo foi modelado por um processo estocástico de ramificação. Para caracterizar a propagação de incertezas na propagação da doença, foram construídos histogramas do número de infectados ao longo do tempo. Também foram calculadas estatísticas amostrais. Foram feitas análises da influência do modelo probabilístico da variável aleatória que modela contágio no processo de ramificação e na probabilidade de extinção da doença ao longo do tempo.

Os resultados do trabalho mostram que os parâmetros do modelo probabilístico da variável aleatória que modela o contágio, C, influenciam fortemente o comportamento a propagação da doença. Sendo C uma $\operatorname{binomial}(m,p)$, o número de indivíduos contaminados por cada membro infectado é limitado a m indivíduos. Como foi visto, para valores de média $\mu \leq 1$ a epidemia se extingue com probabilidade e = 1. Enquanto que, para $\mu > 1$, ela pode se extinguir com probabilidade $e \in [0,1]$. Em trabalhos futuros, seria interessante analisar os efeitos de diferentes modelos probabilísticos de C, como por exemplo v.a. geométrica, na propagação da doença.

AGRADECIMENTOS

Os autores agradecem o apoio dado pela FAPERJ, CNPq e CAPES.

REFERÊNCIAS

- ALLEN, L. An Introduction To Stochastic Processes With Applications To Biology. CRC Press, New York, USA, 2010.
- BRAUER, F., CASTILLO-CHÁVEZ, C., FENG, Z. Mathematical Models In Epidemiology. Springer-Verlag, New York, USA, 2019.
- BRITTON, H., DIECKMANN, T., HEESTERBEEK, O. Mathematical Tools for Understanding Infectious Disease Dynamics. Princeton University Press, Oxford, GBR, 2013.
- GRIMMETT, G., WELSH, D. **Probability An Introduction.** Oxford University Press, Oxford, GBR, 2014.
- HACCOU, P., JAGERS, P., VATUTIN, V. Branching Processes. Cambridge University Press, Cambridge, GBR, 2017.
- Li, M.Y. An Introduction to Mathematical Modeling of Infectious Diseases. Springer, Suiça, 2018.
- LIMA, R., SAMPAIO, R. What is uncertainty quantification? Journal of the Brazilian Society of Mechanical Sciences and Engineering, 40(3), 2018.
- LIMA, R., SAMPAIO, R. How to Deal with Uncertainty Quantification and Propagation. *Mecánica Computacional*, v.XXXVI, 723-739, 2018.
- MARTCHEVA, M. An Introduction To Mathematical Epidemiology. Springer, New York, USA, 2015.
- SAMPAIO, R., LIMA, R. Modelagem Estocástica e Geração de Amostras de Variáveis e Vetores Aleatórios. Notas em Matemática Aplicada; v.70. SBMAC, São Carlos, SP, Brasil, 2012.
- SOUZA de CURSI, J.E., SAMPAIO, R. Modelagem Estocástica e Quantificação de Incertezas. Notas em Matemática Aplicada; v. 66. SBMAC, São Carlos, SP, Brasil, 2012.



How the spread of an infectious disease is affected by the contagion's probabilistic model

Beatriz Borges¹, Roberta Lima¹ e Rubens Sampaio¹

¹ Pontif*ć*ia Universidade Católica do Rio de Janeiro, Rio de Janeiro/RJ, Brasil

Abstract

This work analyzes the spread of an epidemiological disease with a stochastic approach. In the analysis, the number of individuals that each infected member of the population can infect is modeled as a random variable and the number of individuals infected over time is modeled as a stochastic branching process. The focus of the work is to characterize the influence of the probabilistic model of the random variable that models the contagion between individuals in the spread of the disease and in the probability of extinction. The comparison is made based on histograms and sample statistics of the number of infected individuals over time, such as, mean and variance. Statistical models are computed using Monte Carlo simulations to 3 different families of random variables: binomial, geometric-1 and geometric-0. For each family, 21 different distributions were selected and, for each distribution, 2000 simulations of the branching process were computed. In total, 126 thousand simulations were performed, characterizing a big data problem.

Keywords: Stochastic Modeling, Uncertainties Propagation, Infectious Disease, Big Data

1 INTRODUCTION

The spread of infectious diseases is a subject of interest in many different research areas. The knowledge of the spread behavior over time can help governments to plan and take best control strategies. Accurate predictions of the evolution of the number of infected individuals over time can help, for example, in the organization of hospital supplies.

In the literature, there are several different models for the propagation of an infectious disease. Some examples are the compartmental models, as the susceptible-infectious (SI model) and the susceptible-infectious-removed (SIR model). In each of these models, the population is divided into groups (called compartments) and an initial value problem characterizes the evolution of the number of individuals in each of the groups [9], [3], [6].

HOW THE SPREAD OF AN DISEASE IS AFFECTED BY THE CONTAGION'S PROB. MODEL BORGES et al.

Although these models are widely used, they approach the problem of propagation from a deterministic point of view. They do not consider the inherent random behavior of the contagion. For more accurate predictions, stochasticity should be taken into account.

This paper analyzes the spread of an epidemiological disease in a population with a stochastic approach, as found in [2] and [5]. In the analysis, the number of individuals that each infected member of the population can infect is modeled as a random variable and the number of individuals infected over time is modeled as a stochastic branching process. The focus of the work is to characterize the influence of the probabilistic model of the random variable that models the contagion between individuals in the spread of the disease and in the probability of extinction. The comparison is made based on histograms and sample statistics of the number of infected individuals over time, such as, mean and variance. The evolution of the histograms characterizes the uncertainty propagation in the spread of the disease [7], [8]. Statistical models are computed using Monte Carlo simulations to 3 different families of random variables: binomial, geometric-1 and geometric-0. For each family, 21 different distributions were selected and, for each distribution, 2000 simulations of the branching process were computed. In total, 126 thousand simulations were performed and over 2 million realizations of random variables were generated, characterizing a big data problem. The computations were developed in *Matlab* software and consumed more than 2 days of CPU time, in a MacOs system, 1.3 GHz Dual-Core Intel Core i5 processor with 4 GB 1600 MHz DDR3 memory.

CONSTRUCTION OF PROBABILISTIC MODELS TO THE CON- $\mathbf{2}$ TAGION AND SPREAD OF THE DISEASE

In this section, the probabilistic models of the contagion between individuals and the spread of the disease are constructed. The number of individuals that each infected member j of the population can infect is modeled as a random variable, C_j . It is assumed that all C_j are independent and identically distributed (IID). We call C the discrete random variable that models the contagion.

The number of individuals infected over time is modeled as a stochastic branching process I with discrete parameter. Given a set of discrete parameters $N = \{0, 1, 2, 3, \dots\}$, representing the generations over time, for each $n \in N$, $I(n) = I_n$ is a discrete random variable that represents the number of infected individuals in the n^{th} generation. It is given by the sum of the number of individuals who were contaminated by each infected individual of the previous generation (generation n-1), characterizing it as a Markov process, that is:

$$I_n = \sum_{k=0}^{I_{n-1}} C_k, \qquad n > 0.$$
(1)

Figure 1 shows the first five generations of one realization of the branching process I. The process of spread starts with a single infected person $i_0 = 1$. In the next generation, n = 1, 3 new individuals are now infected by the initial one. This amount of new infected individuals is obtained by a sample of the discrete random variable C. In the following generation, n = 2, the total of infected individuals is the sum of the number of individuals who were contaminate by each one of the 3 infected individuals of the generation before. To achieve this number, 3 independent samples of C were generated, and the drawn

PUC-Rio - Certificação Digital Nº 1920913/CA



quantities were added. Given that the first individual infected only one person and, the two others have not infected anyone, $i_2 = 1$. In the next generation, n = 3, the single person infected in generation 2 infects 4 new individuals, thus $i_3 = 4$. In the generation n = 4, the total number of infected individuals, which is the sum of the number of individuals who were contaminated by each of the four infected in n = 3, is obtained generating four independent realizations of C and adding the amounts drawn. Given that, the first infected person infected three individuals, the second infected person did not make any contamination, and the third and fourth persons made two infections each, we have $i_4 = 7$. If none of these seven contaminated individuals make new contaminations, we will have $i_n = 0, \forall n \geq 5$ and, in this case, the disease will be extincted. The probability of extinction of a disease is one of the variables of interest in this work.



Fig. 1: One realization of the branching process that models the number of infected individuals over generations. It is considered $i_0 = 1$.

The probability generating function (pgf) of C is:

$$G_C(x) = \sum_{k=0}^{\infty} p_k x^k, \qquad (2)$$

where each p_k is the probability of C taking the value k. According to theorem 4.36 of [4], the probability generating function (pgf) of I_n is given by:

$$G_{I_n}(x) = G_C(G_{I_{n-1}}(x)).$$
 (3)

3 DISEASE SPREAD STATISTICS

This section shows some statistics of the stochastic process that models the spread of the disease,[1], [2] and [5]. According to theorem 9.8 and the theory of probability generating function from [4], given that $E(I_n) = E(I_{n-1})\mu$, the average of the number of infected individuals in each generation is $E(I_n) = \mu^n$, where $\mu = \sum_k k p_k$ is the mean of the random variable C. Long-term we have:

$$\lim_{n \to \infty} E(I_n) = \begin{cases} 0 & \text{if } \mu < 1, \\ 1 & \text{if } \mu = 1, \\ \infty & \text{if } \mu > 1. \end{cases}$$
(4)

HOW THE SPREAD OF AN DISEASE IS AFFECTED BY THE CONTAGION'S PROB. MODEL BORGES et al.

Adapting theorem 9.8 from [4], it is possible to write the variance of the number of infected individuals in each generation as:

$$var(I_n) = \begin{cases} n\sigma^2 & \text{if } \mu = 1, \\ \sigma^2 \mu^{n-1} \left(\frac{\mu^n - 1}{\mu - 1}\right) & \text{if } \mu \neq 1, \end{cases}$$
(5)

where σ^2 the variance of C. Long-term:

$$\lim_{n \to \infty} var(I_n) = \begin{cases} 0 & \text{if } \mu < 1, \\ \infty & \text{if } \mu \ge 1. \end{cases}$$
(6)

Analyzing Eqs. 4 and 6, we observe that for $\mu \leq 1$ the epidemic will be extinct someday. For $\mu > 1$, on the other hand, the epidemic may or may not end. To determine the probability of extinction, we define two events. First event: A, extinction of the disease in the n^{th} generation, i.e., $I_n = 0$. Evaluating Eq. 3 at x = 0, we get:

$$G_{I_n}(x) = G_C(G_{I_{n-1}}(x))$$

$$G_{I_n}(0) = G_C(G_{I_{n-1}}(0)) = e_n,$$
(7)

where e_n is the probability of A. Second event: B, the disease extinction. By theorem 1.54 from [4], the probability of this event is given by $\lim e_n$, represented by e. Therefore:

$$e_n = G_{I_n}(0) = G_C(G_{I_{n-1}}(0)) \quad \rightarrow \quad e_n = G_C(e_{n-1})$$
$$\lim_{n \to \infty} e_n = e \quad \rightarrow \quad e = G_C(e), \tag{8}$$

where the smallest root of the equation $e = G_C(e)$ in [0,1] is the probability of the epidemic extinction.

4 ESTIMATED STATISTICS, HISTOGRAMS, AND PROBABILITY OF EXTINCTION

To analyze the influence of the probabilistic model of the random variable that models the contagion between individuals in the spread of the disease and and probability of extinction, statistical models were computed for the disease propagation using Monte Carlo simulations to 3 different families of random variables C: binomial, geometric-1 and geometric-0. For each family, 21 different distributions were selected and, for each distribution, 2000 simulations of the branching process from generation n = 0 to n = 20were computed. In total, 126 thousand simulations were performed. A convergence study was developed to determine the number of simulations [11]. A chart of the simulations is shown in Fig.2. In order to analyze the stochastic behavior of the epidemic over the generations, sample statistics, as mean, variance and probability of extinction, and normalized histograms of the number of infected individuals were computed for different generations of the branching process. With those histograms is possible to observe the propagation of uncertainties [10] in the spread of the disease. The most expressive results obtained with the Monte Carlo simulations are shown in the following sections.

Apêndice A. Artigos Publicados

Anais do XIV Encontro Acadêmico de Modelagem Computacional.

Laboratório Nacional de Computação Científica, 08 a 11 de Fevereiro de 2021.



96

Mean, variance and histograms

Fig. 2: Chart of the performed simulations.

4.1 Probabilistic model of C - binomial(m, p)

The first family chosen to C was the binomial(m, p) distribution. This distribution can be interpreted as an experiment in which an individual meets m others individuals and have probability p of infecting each one of them. After those meetings, we evaluate whether each one of the *m* individuals was infected, and count the number of new infected. A binomial(m, p) distribution has mean $\mu = mp$, variance $\sigma^2 = mp(1-p)$ and pgf $G_C(x) = (q + px)^m$. Figure 3 shows the normalized histograms of the number of infected individuals constructed for different generations of the branching process with contagion modeled by the binomial (3, 0.5). Please observe the evolution of the function mass of the number of infected throughout the generations. Figure 4 shows the sample mean and variance of the process I over the generations for three different values of p (p = 0.16, 0.33 and 0.50). The results agree with the analytical results given by Eqs. (4) and (6). For p = 0.16, with $\mu = 0.5$, we can see from Fig. 4(a) that both average and variance of the stochastic process tends to zero, indicating that for $\mu < 1$ we have the extinction of the epidemic. For p = 0.33, with $\mu = 1.0$, we can see from Fig. 4(b) that the process average is around 1, while the variance grows linearly along of the generations. This result is in accordance with Eq. (5). For p = 0.50, with $\mu = 1.5$, we can see from Fig.4(c) that both mean and process' variance grow over the generations. This behavior indicates that $\lim_{n\to\infty} E(I_n) = \infty$ and $\lim_{n\to\infty} var(I_n) = \infty$, so that there is no possibility of epidemic extinction. According to Eq.(8), the probability of extinction of the disease, e, is given by $G_C(e) = (q + pe)^m = e$. Figure 5 shows the probabilities of extinctions over the generations for the three values of p set above. Note that, with probability e = 1, the epidemic tends to extinguish for values of $\mu < 1$, and that for values of $\mu < 1$, it extinguishes with values of $e \in [0, 1]$.



HOW THE SPREAD OF AN DISEASE IS AFFECTED BY THE CONTAGION'S PROB. MODEL BORGES et al.

Fig. 3: Normalized histograms of the number infected individuals in the 5^{th} , 10^{th} , 15^{th} and 20^{th} generations constructed with 2000 realizations of the branching process I, with C = binomial(3, 0.50).



Fig. 4: Sample's and analytic mean and variance for the different values of p.



Fig. 5: Analytic and sample's probability of extinction for the different values of p.

Apêndice A. Artigos Publicados

Anais do XIV Encontro Acadêmico de Modelagem Computacional. Laboratório Nacional de Computação Científica, 08 a 11 de Fevereiro de 2021.



4.2 Probabilistic model of C - geometric-1(p)

The second family chosen to C was the geometric-1(p) distribution. On this less conservative analysis, an infected individual has no limit of individuals that he can infect, and, always, infects at least one person. What is singular in this analysis is that: if every individual infects at least one other individual, this epidemic will never be over. This model depicts a society with low, or not at all, restrictions concerning social distancing. The geometric-1 distribution has its pgf given by $G_C(x) = \frac{px}{1 - (1 - p)x}$, its mean $\mu = \frac{1}{p}$,

and variance $\sigma^2 = \frac{1-p}{p}$. For every value of $p \neq 0$ we have $\mu \geq 1$, in agreement with the fact that the epidemic will never get extinguished. This implies that the probability of extinction is e = 0.



Fig. 6: Normalized histograms of the number infected individuals in the 5^{th} , 10^{th} , 15^{th} and 20^{th} generations constructed with 2000 realizations of the branching process I, with C =geometric-1(0.88).



Fig. 7: Sample's and analytic mean and variance for the different values of p.



HOW THE SPREAD OF AN DISEASE IS AFFECTED BY THE CONTAGION'S PROB. MODEL BORGES et al.

Fig. 8: Analytic and sample's probability of extinction for the different values of p.

Figure 6 shows the normalized histograms of the number of infected individuals constructed for different generations of the branching process with contagion modeled by the geometric-1(p), with p = 0.88. Figure 7 shows the agreement between the sample statistics, mean and variance, and the analytic ones. Figure 8 shows the probability of extinction $e_n = 0$ over the generations.

4.3 Probabilistic model of C - geometric-O(p)

The third family chosen to C was as the geometric-0(p). In this scenario, an infected individual has no limit of individuals that he can infect, but he has the possibility of not infecting anyone. Geometric-0 portrays a society with low restrictions of social distancing, but with some isolated individuals. The pgf of the geometric-0 distribution is $G_C(x) = \frac{p}{1-(1-p)x}$, and its mean is $\mu = \frac{1-p}{p}$. Now, as the individuals can go out and not infect someone, there is a non null probability of extinction of the epidemic. By Eq. 8, with C being geometric-0, the probability of extinction for $\mu > 1$ is given by $e = \frac{p}{1-p}$. Figure 10 shows the agreement between the sample statistics, mean and variance, and the analytic ones. Figure 11 shows the probability of extinction $e_n = 0$ over the generations. Figure 9 shows the normalized histograms of the number of infected individuals constructed for different generations of the branching process with contagion modeled by the geometric-0(p), with p = 0.4.

5 CONCLUSIONS

This work analyzes with a stochastic approach the spread of an epidemiological disease in a population. The number of individuals infected by each infected member of the population was modeled by a discrete random variable and the number of infected individuals over the time was modeled by a stochastic branching process. To characterize the propagation of uncertainties in the spread of the disease, histograms of the number of infected individuals were constructed. Sample statistics were also calculated. Analysis of the influence of the probabilistic model of the random variable that models the contagions between individuals in the branching process were made and the probability of extinction was calculated for three different families of random variables.

The results of the work show that the parameters of the probabilistic model of the random variable that models the contagion, C, strongly influence the propagation's behavior of the disease. C being a binomial(m, p), the number of individuals infected by

Anais do XIV Encontro Acadêmico de Modelagem Computacional.

Laboratório Nacional de Computação Científica, 08 a 11 de Fevereiro de 2021.



100

Fig. 9: Normalized histograms of the number infected individuals in the 5th, 10th, 15th and 20th generations constructed with 2000 realizations of the branching process I, with C = geometric-0(0.4).



Fig. 10: Sample's and analytic mean and variance for the different values of p.



Fig. 11: Analytic and sample's probability of extinction for the different values of p.

each infectious member is limited to m individuals. C being either geometric-1(p) or

HOW THE SPREAD OF AN DISEASE IS AFFECTED BY THE CONTAGION'S PROB. MODEL BORGES et al.

geometric-0(p) there is no limit on how many new individual can be infected. As has been seen, for the binomial(m, p) and geometric-0(p), for values of $\mu \leq 1$ the epidemic is extinguished with probability e = 1. Whereas, for $\mu > 1$, it can be extinguished with probability $e \in [0.1]$. For the geometric-1 distribution, since for every $p \neq 0$ the mean is $\mu > 1$, the probability of extinction is e = 0.

6 Acknowledgments

The authors acknowledge the support given by FAPERJ, CNPq, and CAPES.

References

- L. Allen. An Introduction To Stochastic Processes With Applications To Biology. CRC Press, New York, USA, 2010.
- [2] L. Allen. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. *Infectious Disease Modelling*, 2:128–142, 2017.
- [3] F. Brauer, C. Castillo-Chavez, and Z. Feng. Mathematical Models In Epidemiology. Springer, New York, USA, 2019.
- [4] G. Grimmett and D. Welsh. Probability An Introduction. Oxford University Press, Oxford, GBR, 2014.
- [5] P. Haccou, P. Jagers, and V. Vatutin. *Branching Processes: Variation, Growth, and Extinction of Populations.* Cambridge University Press, Cambridge, GBR, 2017.
- [6] M. Li. An Introduction To Mathematical Modeling of Infectious Diseases. Springer, Switzerland, 2018.
- [7] R. Lima and R. Sampaio. How to deal with uncertainty quantification and propagation. Asociación Argentina de Mecánica Computacional, XXXVI:723–739, 2018.
- [8] R. Lima and R. Sampaio. What is uncertainty quantification? Journal of the Brazilian Society of Mechanical Sciences and Engineering, 40:155, 2018.
- [9] M. Martcheva. An Introduction To Mathematical Epidemiology. Springer, New York, USA, 2015.
- [10] R. Sampaio and R. Lima. Modelagem Estocástica e Geração de Amostras de Variáveis e Vetores Aleatórios, volume 70. Notas de Matemática Aplicada, SBMAC, São Carlos, SP, Brazil, 2012.
- [11] J. E. Souza de Cursi and R. Sampaio. Modelagem Estocástica e Quantificação de Incertezas, volume 66. Notas de Matemática Aplicada, SBMAC, São Carlos, SP, Brazil, 2012.

INSTITUTO PEDERAL



ANÁLISE ESTOCÁSTICA DE PROPAGAÇÃO DE DOENÇAS EPIDEMIOLÓGICAS

STOCHASTIC ANALYSIS OF THE SPREAD OF EPIDEMIOLOGICAL DISEASES

Beatriz Borges¹ Roberta Lima² Rubens Sampaio³

Resumo: Este trabalho analisa com uma abordagem estocástica a propagação de uma doença epidemiológica em uma população. O foco é caracterizar a evolução do número de infectados ao longo do tempo de um ponto de vista probabilístico, ou seja, caracterizar a propagação de incertezas no número de infectados. Para fazer essa análise, o número de indivíduos contaminados por cada membro já infectado da população é modelado como uma variável aleatória binomial(m, p) e a propagação da doença é modelada como um processo estocástico de ramificação. Um modelo estatístico é construído para esse processo de ramificação. Histogramas e estatísticas amostrais do número de infectados ao longo do tempo, como por exemplo, média e variância, são calculadas através de simulações de Monte Carlo para o processo de propagação da doença. Os objetivos do trabalho são comparar as estatísticas amostrais calculadas com previsões da literatura, e determinar a influência de parâmetros do modelo probabilístico da variável aleatória de contágio no processo de ramificação. Através dessas análises determina-se a probabilidade de extinção da doença.

Palavras-chave: Modelagem estocástica. Propagação de incertezas. Processo de ramificação. Epidemiologia.

²Dr., Pontifícia Universidade Católica do Rio de Janeiro, robertalima@puc-rio.br.

¹Bacharel, Pontifícia Universidade Católica do Rio de Janeiro, beatrizrbborges@gmail.com.

³Dr., Pontifícia Universidade Católica do Rio de Janeiro, rsampaio@puc-rio.br.

Abstract: This paper analyzes the spread of an epidemiological disease in a population with a stochastic approach. The focus is to characterize the evolution of the number of infected individuals over time from a probabilistic point of view, that is, to characterize the propagation of uncertainties in the number of infected individuals. To make this analysis, the number of individuals that each infected member of the population can infect is modeled by a binomial random variable with parameters m and p, and the spread of the disease is modeled as a stochastic branching process. A statistical model is built for this branching process. Histograms and sample statistics of the number of infected individuals over time, such as mean and variance, are computed using Monte Carlo simulations. The objectives of the work are to compare the computed sample statistics with some forecasts found in the literature, and to determine the influence of the parameters of the contagion's probabilistic model in the spread of the disease. Through these analysis, it is possible to determine the probability of extinction of the disease.

Keywords: Stochastic modeling. Uncertainty propagation. Branching process. Epidemiology.

1 INTRODUÇÃO

Propagação de doenças epidemiológicas é um tema que vem atraindo interesse de pesquisadores de diferentes áreas. Saber como doenças se propagam, pode facilitar o planejamento e implementação de medidas de controle e contenção. Além disso, prever o número de infectados ao longo do tempo permite que governos coordenem ações e melhor atendam as necessidades da população, por exemplo, adequando o número de leitos hospitalares ao número de infectados que precisam de internação.

Na literatura existem diversos modelos matemáticos para modelar a propagação de doenças, como por exemplo, o modelo de compartimentos de susceptíveis e infectados (SI) e o modelo de compartimentos de susceptíveis, infectados e recuperados (SIR). Em cada um desses modelos, a população é dividida em grupos (chamados de compartimentos) e um problema de valor inicial caracteriza a evolução do número de indivíduos em cada um desses grupos (MARTCHEVA, 2015; BRAUER; CASTILLO-CHAVEZ; FENG, 2019; LI, 2018; DIEKMANN; HEESTERBEEK; BRITTON, 2013).

Apesar de amplamente utilizados em previsões, esses modelos abordam o processo de propagação de doenças de um ponto de vista determinístico, quando na verdade o problema é estocástico. A incerteza no contágio faz com que o número de infectados ao longo do tempo seja aleatório. Para fazermos melhores previsões da propagação, devemos levar em consideração essa estocasticidade.

Este trabalho analisa a propagação de uma doença epidemiológica entre indivíduos de uma população sob uma ótica estocástica. O número de indivíduos contaminados por cada membro já infectado da população é modelado por uma variável aleatória discreta e o número de indivíduos infectados ao longo do tempo é modelado por um processo estocástico de ramificação. O foco do trabalho é caracterizar a propagação de incertezas no número de infectados, ou seja caracterizar como as funções de massa de seções do processo variam ao longo do tempo (LIMA; SAMPAIO, 2018b; LIMA; SAMPAIO, 2018a). Para fazer essa análise, um modelo estatístico é construído para esse processo de ramificação. Histogramas e estatísticas amostrais do número de infectados ao longo do tempo, como por exemplo, média e variância, são calculadas através de simulações de Monte Carlo para o processo de propagação da doença. As simulações são feitas através do software *Matlab*. Os objetivos do trabalho são comparar as estatísticas amostrais calculadas com previsões da literatura, e determinar a influência de parâmetros do modelo probabilístico da variável aleatória de contágio no processo de ramificação. Através dessas análises deseja-se determinar qual é a probabilidade de extinção da doença (BORGES; LIMA; SAMPAIO, 2020; BORGES; LIMA; SAMPAIO, 2021).

2 CONSTRUÇÃO DE MODELOS PROBABILÍSTICOS PARA O CONTÁGIO E PROPAGAÇÃO DA DOENÇA

Esta seção trata da modelagem probabilística do contágio entre indivíduos da população e da propagação da doença. O número de indivíduos contaminados por cada membro já infectado j da população é modelado como uma variável aleatória (v.a.) discreta binomial(m, p) denominada C_j . Supõe-se que as v.a. C_j sejam independentes e identicamente distribuídas (IID). Com isso, denominamos C a variável aleatória que modela o contágio. A distribuição binomial pode ser interpretada, de acordo com conceitos de epidemiologia, como um experimento no qual um indivíduo encontra com um número m de pessoas, com uma probabilidade p de infectar cada uma delas. Depois dos encontros, avaliamos se cada uma das m pessoas foi, ou não, infectada e contamos o número de novos contagiados.

O número de infectados ao longo do tempo é modelado por um processo estocástico I de parâmetro discreto chamado de processo de ramificação. Dado um conjunto de parâmetros discretos $N = \{0, 1, 2, 3, ...\}$, representando gerações ao longo do tempo, para cada $n \in N$, $I(n) = I_n$ é uma variável aleatória discreta. I_n representa o número de infectados na n-ésima geração e é dado pela soma do número de pessoas que foram contagiadas por cada um dos infectados na geração anterior n - 1, ou seja:

$$I_n = \sum_{k=0}^{I_{n-1}} C_k , \qquad n > 0 .$$
 (1)

A figura 1 ilustra as cinco primeiras gerações de uma realização do processo de ramificação I. Nesta realização, o processo de contaminação começa com um indivíduo infectado, $i_0 = 1$. Na geração n = 1, aparecem 3 indivíduos contaminados pelo indivíduo infectado inicialmente, $i_1 = 3$. Esse número de infectados é obtido fazendo-se uma realização da variável aleatória C. Na próxima geração, n = 2, o número total de infectados é a soma do número de pessoas que foram contagiadas por cada um dos três infectados na geração anterior. Para obter esse número, geramos três realizações independentes de C e somamos os valores sorteados. No exemplo ilustrado, o primeiro infectado contaminou apenas um indivíduo e os outros dois não fizeram nenhuma contaminação. Dessa forma $i_2 = 1$. Na próxima geração, n = 3, o único indivíduo contaminado na geração anterior contaminou quatro novos indivíduos, fazendo com que $i_3 = 4$. Na geração n = 4, o número total de infectados, que é a soma do número de pessoas que foram contagiadas por cada um dos quatro infectados em n = 3, é obtido gerando-se quatro realizações independentes de C e somando os valores sorteados. No exemplo ilustrado, o primeiro infectado contaminou dois indivíduos, o segundo infectado não fiz nenhuma contaminação e o terceiro e quarto fizeram duas contaminações cada. Totalizando $i_4 = 7$. Se nenhum desses sete indivíduos contaminados fizer novas contaminações, teremos $i_n = 0, \, \forall n \geq 5$, e nesse caso a doença estará extinta. A probabilidade de extinção de uma doença é uma das variáveis de interesse desse trabalho.

A função geradora de probabilidade de C é:

$$G_C(x) = \sum_{k=0}^{\infty} p_k x^k , \qquad (2)$$

onde p_k é a probabilidade de *C* assumir o valor *k*. Sendo *C* uma a variável aleatória binomial(m, p), a função geradora de probabilidade de *C* pode ser reescrita como:

$$G_C(x) = ((1-p) + px)^m .$$
(3)

De acordo com o teorema 4.36 de Grimmet e Welsh (2014), a função geradora de probabilidade de I_n é dada por:

$$G_{I_n}(x) = G_C(G_{I_{n-1}}(x)).$$
(4)

4

Figura 1: Cinco primeiras gerações de uma realização do processo de ramificação I. Considera-se $i_0 = 1$.



Fonte: (GRIMMETT; WELSH, 2014)

ESTATÍSTICAS DA PROPAGAÇÃO DA DOENÇA 3 PUC-Rio - Certificação Digital Nº 1920913/CA

Esta seção mostra algumas estatísticas do processo estocástico que modela a propagação da doença (HACCOU et al., 2005; ALLEN, 2010). De acordo com o teorema 9.8 de de Grimmet e Welsh (2014), a média de infectados ao longo das gerações é:

$$E(I_n) = \mu^n , \qquad (5)$$

onde $\mu = mp$ é a média da variável aleatória C, binomial(m, p). A longo prazo:

$$\lim_{n \to \infty} E(I_n) \to \begin{cases} 0, & \text{se } \mu < 1, \\ 1, & \text{se } \mu = 1, \\ \infty, & \text{se } \mu > 1. \end{cases}$$
(6)

Fazendo uma adaptação do teorema 9.8 de Grimmet e Welsh (2014), pode-se escrever:

$$var(I_n) = \begin{cases} n\sigma^2, & \text{se } \mu = 1, \\ \sigma^2 \mu^{n-1}\left(\frac{\mu^n - 1}{\mu - 1}\right), & \text{se } \mu \neq 1. \end{cases}$$
(7)

onde $\sigma^2 = mp(1-p)$ é a variância de C. Analisando a variância a longo prazo:

$$\lim_{n \to \infty} var(I_n) \to \begin{cases} 0, & \text{se } \mu < 1, \\ \infty, & \text{se } \mu \ge 1. \end{cases}$$
(8)

A partir da análise mostrada nas equações (6) e (8), temos que se $\mu \leq 1$ a epidemia, eventualmente, chegará ao fim. Para $\mu > 1$, por sua vez, a epidemia pode, ou não, chegar ao fim. Para determinar essa probabilidade de extinção definimos dois eventos distintos. Primeiro definimos um evento A: extinção da doença na *n*-ésima geração, i.e., $I_n = 0$. Utilizando a equação (4) e avaliando-a em x = 0 temos:

$$G_{I_n}(x) = G_C(G_{I_{n-1}}(x))$$

$$G_{I_n}(0) = G_C(G_{I_{n-1}}(0)) = e_n ,$$
(9)

assim, chamamos de e_n a probabilidade de A. Depois, definimos o evento B: extinção da doença. De acordo com o teorema 1.54 de Grimmet e Welsh (2014), a probabilidade deste evento ocorrer é dada pelo $\lim_{n\to\infty} e_n$ que será representado por e. Assim:

$$e_n = G_{I_n}(0) = G_C(G_{I_{n-1}}(0)) \quad \rightarrow \qquad e_n = G_C(e_{n-1})$$
$$\lim_{n \to \infty} e_n = e \quad \rightarrow \qquad e = G_C(e) , \tag{10}$$

onde a menor raiz da equação $e = G_C(e)$ no intervalo [0, 1] é a probabilidade da epidemia se extinguir.

4 ESTATÍSTICAS AMOSTRAIS, PROBABILIDADE DE EXTINÇÃO E HISTOGRAMAS NORMALIZADOS

Para construir um modelo estatístico por Monte Carlo para a propagação da doença de forma poder compará-lo com o modelo encontrado na literatura, mostrado na seção 3, foram feitas 2,000 realizações de I da geração n = 0 até n = 20 através do software *Matlab*. Um estudo de convergência foi desenvolvido para determinar o número de realizações utilizadas (SOUZA DE CURSI; SAMPAIO, 2012). Estatísticas amostrais (média, variância e probabilidade de extinção) e histogramas foram construídos com as realizações.

Com o objetivo de analisar o comportamento estocástico da epidemia ao longo das gerações, foram construídos os histogramas normalizados do número de infectados para diferentes gerações do processo de ramificação com o contagio modelado pela binomial(3, 0.5). Os resultados estão mostrados na figura 2. Nessa figura, é possível observar a propagação de incertezas (SAMPAIO; LIMA, 2012) na propagação da doença, i.e., a evolução da função de massa do número de infectados ao longo das gerações.

A figura 3 mostra a média e variância amostrais do processo I ao longo das gerações para diferentes valores de p. Considera-se p = 0.16, 0.33 e 0.50. Observamos que os valores amostrais estão de acordo com os resultados analíticos dados pelas equações (6) e (8).

Para p = 0.16, com $\mu = 0.5 < 1$, podemos observar pela figura 4(a) que tanto a média quanto a variância do processo estocástico tendem a zero, confirmando que para $\mu < 1$ temos a extinção da epidemia. Para p = 0.33, com $\mu = 1.0$, podemos observar pela figura 4(b) que a média do processo fica em torno de 1, enquanto a variância cresce linearmente ao longo das gerações. Esse resultado está de acordo com a equação (7). Para p = 0.50, com $\mu = 1.5 > 1$, podemos observar pela figura 4(c) que tanto a média quanto a variância do processo crescem ao longo das gerações. Esse comportamento indica que $\lim_{n\to\infty} E(I_n) = \infty$ e $\lim_{n\to\infty} var(I_n) = \infty$, de forma que não há possibilidade da epidemia se extinguir.

De acordo com a equação (10), a probabilidade de extinção da doença, e, é dada por $G_C(e) = ((1 - p) + pe)^m = e$. A figura 4 mostra as probabilidades de extinção amostrais

Figura 2: Histogramas normalizados do número de infectados nas gerações 5, 10, 15 e 20, construídos com 2,000 realizações da binomial(3, 0.5).



Fonte: Do autor

ao longo das gerações para os diferentes valores de p fixados acima. Observe que, com probabilidade e = 1, a epidemia tende a se extinguir para valores de $\mu \le 1$, e que, para valores de $\mu > 1$, ela se extingue com valores de $e \in [0, 1]$.
Figura 3: Média e variância, analíticas e amostrais, para os valores p: 0.16, 0.33 e 0.50.



Fonte: Do autor

Figura 4: Gráfico da probabilidade de extinção da doença, analítica e amostral para os diferentes valores p: 0.16, 0.33 e 0.50.



5 CONCLUSÕES

Este trabalho utilizou uma abordagem estocástica para modelar o problema de propagação de uma doença epidemiológica em uma população. O número de indivíduos contaminados por cada membro infectado da população foi modelado por uma variável aleatória (v.a.) discreta binomial(m, p) e o número de infectados ao longo do tempo foi modelado por um processo estocástico de ramificação. Para caracterizar a propagação de incertezas na propagação da doença, foram construídos histogramas do número de infectados ao longo do tempo. Também foram calculadas estatísticas amostrais. Foram feitas análises da influência de parâmetros do modelo probabilístico da variável aleatória que modela contágio no processo de ramificação e na probabilidade de extinção da doença.

Os resultados do trabalho mostram que os parâmetros do modelo probabilístico da variável aleatória que modela o contágio, C, influenciam fortemente o comportamento a propagação da doença. Sendo C uma binomial(m, p), o número de indivíduos contaminados por cada membro infectado é limitado a m indivíduos. Como foi visto, para valores de média $\mu \leq 1$ a epidemia se extingue com probabilidade e = 1. Enquanto que, para $\mu > 1$, ela pode se extinguir com probabilidade $e \in [0, 1]$. Em trabalhos futuros, seria interessante analisar os efeitos de diferentes modelos probabilísticos de C, como por exemplo v.a. geométrica, na propagação da doença.

Referências

ALLEN, L. An Introduction to Stochastic Processes with Applications to Biology, Second Edition. New York, USA: Taylor & Francis, 2010. 490 p. ISBN 9781439818824. 5

BORGES, B.; LIMA, R.; SAMPAIO, R. Análise estocástica de propagação de doenças epidemiológicas. In: Anais do X Encontro Regional de Matemática Aplicada e Computacional do Rio Grande do Sul – ERMAC-RS. Brasil: EDIPUCRS, 2020. p. 8. Disponível em: https://ebooks.pucrs.br/edipucrs/anais/1501/assets/edicoes/2020/arquivos/83. pdf>. 3

BORGES, B.; LIMA, R.; SAMPAIO, R. How the spread of an infectious disease is affected by the contagion's probabilistic model. In: **Anais do XIV Encontro Acadêmico de Modelagem Computacional.** Brasil: LNCC, 2021. p. 6–16. Disponível em: http://www.eamc.lncc.br/PastEditions/Proceedings/Proceedings_EAMC2021.pdf>. 3

BRAUER, F.; CASTILLO-CHAVEZ, C.; FENG, Z. Mathematical Models in Epidemiology. New York, USA: Springer, 2019. 625 p. (Texts in Applied Mathematics). ISBN 9781493998289. 3

DIEKMANN, O.; HEESTERBEEK, H.; BRITTON, T. Mathematical Tools for Understanding Infectious Disease Dynamics. Oxford, GBR: Princeton University Press, 2013. 981 p. (EBSCO ebook academic collection). ISBN 9780691155395. 3

GRIMMETT, G.; WELSH, D. **Probability: An Introduction**. Oxford, GBR: Oxford University Press, 2014. 281 p. (Oxford University Press). ISBN 9780198709978. 4, 5, 6

HACCOU, P. et al. **Branching Processes: Variation, Growth, and Extinction of Populations**. Cambridge, GBR: Cambridge University Press, 2005. 330 p. (Cambridge Studies in Adaptive Dynamics). ISBN 9780521832205. 5

LI, M. **An Introduction to Mathematical Modeling of Infectious Diseases**. [S.l.]: Springer International Publishing, 2018. 156 p. (Mathematics of Planet Earth). ISBN 9783319721224. 3

LIMA, R.; SAMPAIO, R. How to deal with uncertainty quantification and propagation. **Mecánica Computacional**, XXXVI, n. 16, p. 723–739, 2018. Disponível em: https://cimec.org.ar/ojs/index.php/mc/article/view/5572/5550>. 3

LIMA, R.; SAMPAIO, R. What is uncertainty quantification? Journal of the Brazilian Society of Mechanical Sciences and Engineering, v. 40, n. 3, p. 155, 2018. Disponível em: https://doi.org/10.1007/s40430-018-1079-7>. 3

MARTCHEVA, M. An Introduction to Mathematical Epidemiology. New York, USA: Springer US, 2015. 462 p. (Texts in Applied Mathematics). ISBN 9781489976123. 3

SAMPAIO, R.; LIMA, R. Modelagem Estocástica e Geração de Amostras de Variáveis e Vetores Aleatórios. São Carlos, Brasil: SBMAC, 2012. 144 p. (Notas em Matemática Aplicada, vol. 70). ISSN 2236-5915. Disponível em: https://proceedings.science/series/23/proceedings_non_indexed/75. 6

SOUZA DE CURSI, J.; SAMPAIO, R. **Modelagem Estocástica e Quantificação de Incertezas**. SBMAC, 2012. 344 p. (Notas em Matemática Aplicada, vol. 66). ISSN 2175-3385. Disponível em: https://proceedings.science/series/23/proceedings_non_indexed/71.

Edição -

Editor -