

Gabriela Santos Correia

Processo de Ramificação em Modelos Epidemiológicos

Niterói - RJ, Brasil

12 de Dezembro de 2022

Gabriela Santos Correia

Processo de Ramificação em Modelos Epidemiológicos

Trabalho de Conclusão de Curso

Monografia apresentada para obtenção do grau de Bacharel em Estatística pela Universidade Federal Fluminense.

Orientadora: Prof^ª. Dra. Karina Yuriko Yaginuma

Niterói - RJ, Brasil

12 de Dezembro de 2022

Gabriela Santos Correia

**Processo de Ramificação em Modelos
Epidemiológicos**

Monografia de Projeto Final de Graduação sob o título “*Processo de Ramificação em Modelos Epidemiológicos*”, defendida por Gabriela Santos Correia e aprovada em 12 de Dezembro de 2022, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

Profa. Dra. Karina Yuriko Yaginuma
Departamento de Estatística – UFF

Prof. Dr. Jaime Antonio Utria Valdes
Departamento de Estatística – UFF

Prof. Dr. Douglas Rodrigues Pinto
Departamento de Estatística – UFF

Niterói, 12 de Dezembro de 2022

Ficha catalográfica automática - SDC/BIME
Gerada com informações fornecidas pelo autor

C824p Correia, Gabriela Santos
Processo de ramificação em modelos epidemiológicos /
Gabriela Santos Correia. - 2022.
39 f.: il.

Orientador: Karina YuriKo Yaginuma.
Trabalho de Conclusão de Curso (graduação)-Universidade
Federal Fluminense, Instituto de Matemática e Estatística,
Niterói, 2022.

1. Processos estocásticos. 2. Modelos epidemiológicos. 3.
Probabilidade de extinção. 4. Função geradora de
probabilidade. 5. Produção intelectual. I. Yaginuma, Karina
Yuriko, orientadora. II. Universidade Federal Fluminense.
Instituto de Matemática e Estatística. III. Título.

CDD - XXX

Resumo

Os modelos estatísticos desempenham um papel fundamental na vigilância epidemiológica, uma vez que o conhecimento da propagação de uma doença infecciosa facilita o direcionamento de políticas públicas. O processo de ramificação aparece na literatura como uma forma eficaz de entender a propagação de uma epidemia com características similares à de COVID-19, além de apresentar a vantagem da incorporação da aleatoriedade, intrínseca a um processo de contágio, de forma simples. Neste trabalho objetivou-se o estudo de teoria, aplicações e métodos do processo de ramificação. Para isso, desenvolveu-se o funcionamento deste processo, bem como sua média e variância, além do conceito de probabilidade de extinção. É exposto também um estudo de simulação, onde comparam-se os resultados obtidos com aqueles encontrados na literatura.

Palavras-chave: Processos estocásticos. Modelos epidemiológicos. Probabilidade de extinção. Função geradora de probabilidade.

Sumário

Lista de Figuras

1	Introdução	p. 7
1.1	Objetivos	p. 8
2	Materiais e Métodos	p. 9
2.1	O Processo de Ramificação	p. 9
2.2	Esperança e Variância do Processo Z_n	p. 12
2.3	Probabilidade de Extinção	p. 16
3	Análise dos Resultados	p. 22
3.1	Binomial	p. 22
3.2	Geométrica	p. 24
3.3	Poisson	p. 26
4	Conclusão	p. 30
	Referências	p. 32
	Apêndice 1	p. 33
1.1	Momentos de uma variável aleatória discreta	p. 33
1.2	Função Geradora de Probabilidade	p. 34
1.2.1	Definição	p. 34
1.2.2	Momentos	p. 35

Lista de Figuras

1	Um exemplo de processo de ramificação, com gerações de tamanho Z_n (GRIMMET; WELSH, 2014).	p. 10
2	Gráficos da curva $y = G(x)$ com a linha $y = x$ em dois possíveis cenários: à esquerda, com $G'(1) > 1$ e à direita, $G'(1) \leq 1$ (GRIMMET; WELSH, 2014).	p. 21
3	Gráficos das médias teórica e amostral para C com distribuição Binomial de diferentes parâmetros ao longo das gerações.	p. 23
4	Gráficos das variâncias teórica e amostral para C com distribuição Binomial de diferentes parâmetros ao longo das gerações.	p. 24
5	Gráficos das probabilidades de extinção na n -ésima geração teórica e amostral para C com distribuição Binomial de diferentes parâmetros.	p. 24
6	Gráficos das médias teórica e amostral para C com distribuição Geométrica de diferentes parâmetros ao longo das gerações.	p. 25
7	Gráficos das variâncias teórica e amostral para C com distribuição Geométrica de diferentes parâmetros ao longo das gerações.	p. 26
8	Gráficos das probabilidades de extinção na n -ésima geração teórica e amostral para C com distribuição Geométrica de diferentes parâmetros.	p. 27
9	Gráficos das médias teórica e amostral para C com distribuição Poisson de diferentes parâmetros ao longo das gerações.	p. 27
10	Gráficos das variâncias teórica e amostral para C com distribuição Poisson de diferentes parâmetros ao longo das gerações.	p. 28
11	Gráficos das probabilidades de extinção na n -ésima geração teórica e amostral para C com distribuição Poisson de diferentes parâmetros.	p. 29

1 Introdução

A COVID-19 é uma doença infecciosa, causada pelo vírus SARS-CoV-2, transmitida principalmente pelo contato direto ou indireto com pequenas partículas líquidas excretadas pelo nariz ou boca de pessoas infectadas (ORGANIZATION, 2020). Sua chegada em 2019 na forma de casos isolados, rapidamente evoluiu para um estado de epidemia, fazendo com que a incerteza sobre o comportamento da propagação da doença afetasse a todos. Neste contexto, ferramentas capazes de delinear o comportamento a princípio desconhecido e imprevisível de uma doença letal e altamente contagiosa como esta (MOHAPATRA et al., 2020) são cruciais para a proteção da população.

Os modelos epidemiológicos possibilitam, através da estimativa de seus parâmetros, projetar a evolução do impacto de uma epidemia como a da COVID-19. Estes são inicialmente estimados utilizando dados relativos a infecções causadas pelo vírus da epidemia em outras regiões ou épocas (FRANCO; DUTRA, 2021). Um parâmetro muito utilizado nestes modelos é a média de indivíduos contaminados deixados por um único indivíduo doente μ . Este valor fornece o resultado expressivo de que a condição de eliminação da doença será garantida quando μ é mantido abaixo de um (FARRINGTON; KANAAN; GAY, 2003). O monitoramento deste valor também está fortemente associado às medidas de vigilância mais efetivas no que se refere ao contágio de uma doença infecciosa com o mesmo processo de contaminação da COVID-19, como programas de vacinação em massa (FARRINGTON; KANAAN; GAY, 2003) e políticas de distanciamento social (FRANCO; DUTRA, 2021)(ZHAO et al., 2021).

O processo de ramificação é um processo estocástico que desempenha um papel importante no entendimento de contágio epidêmico. Este, em contraste com outros modelos epidemiológicos de grande destaque, é alimentado por dados somente sobre os casos da doença, o que o caracteriza como um modelo simples e apropriado para os estágios iniciais de uma epidemia (FARRINGTON; KANAAN; GAY, 2003). Uma outra particularidade deste tipo de modelo que o difere de outros como o Suscetível-Infectado (SI) e Suscetível-Infectado-Recuperado (SIR), amplamente utilizados, é que aproxima a descrição

do contágio à realidade, através da integração de um traço tão presente na propagação de uma doença infecciosa, a aleatoriedade da contaminação. A dependência em modelos estocásticos, por sua vez, os tornam mais complexos que os determinísticos, e, por vezes, sua aplicação é inviável. Contudo, a natureza desse processo fornece muitas vantagens, tornando-o preferível quando é possível utilizá-lo para análises (ANDERSSON; BRITTON, 2000).

É relevante, então, a existência de um modelo simples e eficaz no contexto de surgimento de epidemia, marcado por urgência e falta de informações. Paralelamente, a modelagem da infecção por COVID-19 através de um processo de ramificação mostra-se apropriada devido ao seu traço infeccioso e modo de contágio (ANDERSSON; BRITTON, 2000). Em outros termos, epidemias como a da COVID-19, quando em estágios iniciais, podem se beneficiar expressivamente de um modelo de ramificação.

(BORGES, 2021) apresenta uma análise estocástica epidemiológica, através de estudo de simulação, modelando a propagação de uma doença por meio do processo de ramificação, onde a quantidade de indivíduos infectados por um único indivíduo contaminado é uma variável aleatória que segue a distribuição de probabilidade Binomial. Complementarmente, a proposta do trabalho aqui exposto é realizar o mesmo tipo de estudo, extendendo as análises a outras formas de contágio, ou seja, modelando o número de infectados por meio de outras distribuições de probabilidade.

1.1 **Objetivos**

O propósito inicial deste trabalho foi estudar a teoria do processo de ramificação, bem como a importância das suas aplicações em propagação de doenças contagiosas e os métodos que possibilitam essas utilizações. Em um segundo momento, teve-se como objetivo um estudo de simulação, de modo que o número de infectados por um indivíduo contaminado foi estabelecido utilizando as distribuições Binomial, Poisson e Geométrica para, então, verificar a concordância entre as aplicações do modelo epidemiológico de ramificação obtidas e os resultados presentes na literatura sobre este processo.

2 Materiais e Métodos

Neste Capítulo serão definidos os processos de ramificação e suas propriedades.

2.1 O Processo de Ramificação

Esse processo tem origem da ideia de uma cadeia reprodutiva que se comporta da seguinte forma: na geração $n = 0$, há um único indivíduo capaz de se reproduzir; na geração seguinte, $n = 1$, ele morre, tendo deixado uma família de C descendentes, com mesmas capacidades reprodutivas e tempo de sobrevivência. Então, o processo se repete para as gerações seguintes.

Para viabilizar as aplicações do processo, é incorporado o elemento aleatório que caracteriza esse tipo de modelo, de forma que o número de descendentes deixados por um único indivíduo é tratado como uma variável aleatória. Assim, é possível calcular o tamanho das proles de cada um dos indivíduos, o que é suficiente para obter as relações matemáticas desenvolvidas ao longo desse capítulo, que permitem conhecer o comportamento do crescimento dessa população.

No estudo, o contexto não será reprodutivo, mas de infecção, sendo a quantidade de infectados deixados por um único indivíduo infectado a variável aleatória C , tal que

$$\mathbb{P}(C = k) = p_k, \quad k = 0, 1, 2, \dots \quad (2.1)$$

O número total de infectados na geração n , é dado pela soma de indivíduos que foram infectados na geração anterior, e é representado por $Z(n) = Z_n$.

$$Z_n = \sum_{i=1}^{Z_{n-1}} C_i^{n-1}, \quad n = 0, 1, 2, \dots \quad (2.2)$$

Para Z_0 , o total da geração será sempre 1, por definição.

Ademais, os requisitos para esse processo são de que a quantidade de indivíduos que

cada pessoa infecta são variáveis aleatórias independentes e identicamente distribuídas, sendo C_j^n o número de infectados pelo indivíduo j na geração n , com função de probabilidade dada por p_k , $k = 0, 1, \dots$. Além disso, assume-se $p_k \neq 1$, para todo k .

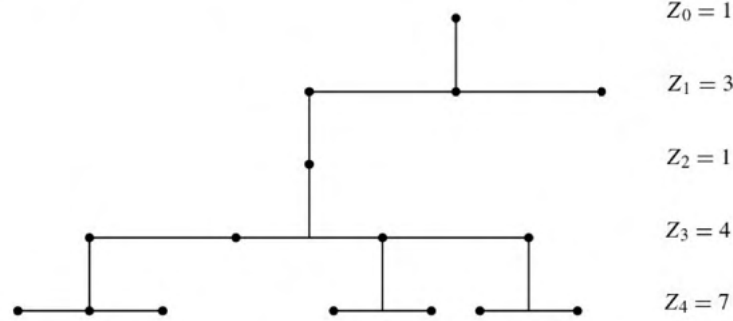


Figura 1: Um exemplo de processo de ramificação, com gerações de tamanho Z_n (GRIMMET; WELSH, 2014).

A Figura 1 apresenta um exemplo de processo de ramificação, que se refere somente aos indivíduos infectados da população. Nela é possível ver que o primeiro indivíduo do processo deixou um total de $C_1^0 = 3$ infectados para a geração $n = 1$ e, por isso, $Z_1 = 3$, o total de contaminados desta geração. Para $n = 2$, o total de indivíduos é a soma dos infectados por cada um dos $k = 3$ indivíduos da geração anterior, isto é,

$$\begin{aligned} Z_2 &= C_1^1 + C_2^1 + C_3^1 \\ &= 1 + 0 + 0 = 1. \end{aligned}$$

Assim segue até a última geração do exemplo, $n = 4$, cujo total de indivíduos, que corresponde à soma de indivíduos que foram infectados por cada membro da geração anterior, pode ser calculado aplicando a fórmula em (2.2)

$$\begin{aligned} Z_4 &= \sum_{i=1}^{Z_3=4} C_i^{m-1} \\ &= C_1^3 + C_2^3 + C_3^3 + C_4^3 \\ &= 3 + 0 + 2 + 2 = 7. \end{aligned}$$

Para encontrar as distribuições de Z_i , considere-se p_k a função de probabilidade do número de infectados por um indivíduo. Pela definição do processo de ramificação utilizado no estudo, $Z_0 = 1$, e

$$\mathbb{P}(Z_1 = k) = p_k, \quad \text{para } k = 0, 1, 2, \dots \quad (2.3)$$

Dessa forma, Z_2 é a soma de contaminados deixados por cada infectado que compôs Z_1 ,

ou seja

$$Z_2 = C_1^1 + C_2^1 + \dots + C_{Z_1}^1.$$

De maneira geral, para qualquer geração n , $n = 1, 2, 3, \dots$, Z_n pode ser escrito como

$$Z_n = C_1^{n-1} + C_2^{n-1} + \dots + C_{Z_{n-1}}^{n-1}. \quad (2.4)$$

A definição da função geradora de probabilidade, presente no Apêndice 1, fornece uma outra forma de representar a soma de uma quantidade aleatória de variáveis aleatórias. No caso de Z_n , a função geradora de probabilidade é dada por

$$G_n(s) = \mathbb{E}(s^{Z_n}) = \sum_{k=0}^{\infty} s^k \mathbb{P}(Z_n = k).$$

E para o tamanho de uma família C gerada por um dos descendentes de qualquer geração, a função geradora de probabilidade é escrita como

$$G(s) = \sum_{k=0}^{\infty} s^k p_k.$$

Do Teorema 1.15, enunciado no Apêndice 1, é possível concluir que a soma Z_n na equação (2.4) pode ser escrita através da função geradora de probabilidade do passo anterior, aplicada à função geradora de probabilidade de C . Ou seja,

$$\begin{aligned} G_0(s) &= s \\ G_n(s) &= G_{n-1}(G(s)), \quad \text{para } n = 1, 2, \dots \text{ ou } n = 2, 3, \dots \end{aligned} \quad (2.5)$$

O resultado em (2.5) é de grande importância para o estudo, pois, a partir dele, derivam outros a respeito da média e variância do processo de ramificação Z_n , desenvolvidas nesta seção. Essas, por sua vez, viabilizam as análises propostas no Capítulo 1.1.

Para mais, há uma outra conclusão a respeito de G_n que mostra ser possível, partindo de (2.5), escrevê-la em função de G , de forma recursiva. Ou seja,

$$G_n(s) = G(G(G(\dots(s)\dots))), \quad \text{para } n = 1, 2, \dots \quad (2.6)$$

Ele deriva da aplicação à G_{n-1} , o resultado obtido anteriormente, e assim sucessivamente

$$\begin{aligned} G_n(s) &= G_{n-2}(G(G(s))) = \dots \\ &= G_1(G(G(\dots(s)\dots))). \end{aligned}$$

Recordando que $G_1 = G$, uma vez que Z_1 tem somente os descendentes de um único indivíduo. Este resultado também será importante para auxiliar no desenvolvimento do estudo, contudo, seu uso ficará reservado à seção seguinte.

2.2 Esperança e Variância do Processo Z_n

Nesta seção, retoma-se o interesse em analisar a distribuição de Z_n , e aprofunda-se o estudo à respeito de sua esperança e variância.

A esperança de Z_n é definida no teorema enunciado abaixo, que trata-se de uma aplicação direta da equação (2.5).

Teorema 2.1. *A esperança de Z_n é definida como*

$$\mathbb{E}(Z_n) = \mu^n,$$

sendo $\mu = \sum_k kp_k$ a média da distribuição do número de infectados.

Demonstração. Derivando o resultado obtido em (2.5), nesta seção, e igualando $s = 1$, obtém-se

$$G'_n(1) = G'_{n-1}(G(1))G'(1).$$

De (1.11), no Apêndice 1, a esperança de Z_n pode ser escrita como $\mathbb{E}(Z_n) = G'_n(1)$. Portanto,

$$\mathbb{E}(Z_n) = G'_{n-1}(G(1))G'(1).$$

O resultado (1.6), no Apêndice 1, fornece $G(1) = 1$, então

$$\mathbb{E}(Z_n) = G'_{n-1}(1)G'(1).$$

De (1.11), do Apêndice 1, também é possível escrever $G'(1) = \mathbb{E}(C)$ — sendo a esperança de C denotada por μ — e $G'_{n-1}(1) = \mathbb{E}(Z_{n-1})$.

Portanto, resulta que

$$\mathbb{E}(Z_n) = \mathbb{E}(Z_{n-1})\mu.$$

Sucessivamente,

$$\begin{aligned}\mathbb{E}(Z_n) &= \mu\mathbb{E}(Z_{n-1}) = \mu^2\mathbb{E}(Z_{n-2}) = \dots \\ &= \mu^n\mathbb{E}(Z_0) = \mu^n.\end{aligned}$$

□

Agora, é possível representar a média de uma variável aleatória cuja distribuição de probabilidade é desconhecida, Z_n , em função da média da distribuição do tamanho de uma família, esta conhecida. A ideia é aplicar essa lógica à variância, igualmente, e escrevê-la em termos de μ e σ^2 , média e variância de C , respectivamente.

Mas, diferentemente da média, a variância de Z_n não é consequência direta da equação (2.5). Então, serão mostrados, primeiramente, outros resultados que auxiliarão no desenvolvimento da definição da variância.

Acima, mostrou-se que da equação (1.11) é possível escrever a primeira derivada de G aplicada a 1 em termos de μ

$$G'(1) = \mu. \quad (2.7)$$

Partindo do resultado (1.12), do Apêndice 1, é possível aplicar a mesma lógica a segunda derivada de G , para $s = 1$, e escrevê-la em termos de μ e também σ^2

$$\begin{aligned}G''(1) &= \mathbb{E}(C^2) - G'(1) && \text{de (1.12)} \\ &= \mathbb{E}(C^2) - \mathbb{E}(C) && \text{de (2.7)}.\end{aligned}$$

De (1.10), no Apêndice 1, é obtida a relação $\sigma^2 = \mathbb{E}(C^2) - E(C)^2$, portanto

$$\begin{aligned}G''(1) &= \sigma^2 + \mathbb{E}(C)^2 - \mathbb{E}(C) \\ &= \sigma^2 + \mu^2 - \mu \\ &= \sigma^2 + \mu(\mu - 1).\end{aligned} \quad (2.8)$$

A partir da derivação de (2.5), resulta, pela regra da cadeia, que a derivada de G_n pode ser escrita em função de μ e da derivada de G_n do passo anterior, quando aplicadas

a 1. Isto é

$$\begin{aligned} G_n(s) &= G_{n-1}(G(s)) \\ G'_n(s) &= G'_{n-1}(G(s))G'(s). \end{aligned} \quad (2.9)$$

De (1.6) no Apêndice, $G(1) = 1$, segue que

$$G'_n(1) = G'_{n-1}(1)\mu. \quad (2.10)$$

Com o auxílio desses resultados, e de outros descritos no Apêndice 1, é possível, enfim, provar que a variância de Z_n pode ser escrita em termos da média e variância de C .

Teorema 2.2. *A variância do tamanho da n -ésima geração de um processo de ramificação, Z_n , é definida como*

$$\text{Var}(Z_n) = \begin{cases} \sigma^2 \mu^{n-1} \left[\frac{\mu^n - 1}{\mu - 1} \right] & \text{se } \mu \neq 1 \\ n\sigma^2 & \text{se } \mu = 1, \end{cases}$$

onde μ é a média da distribuição do tamanho de uma família, σ^2 , a variância, e n , a geração do processo.

Demonstração. Considere a segunda derivada de G_n das duas seguintes formas:

1^a. Desenvolvendo o resultado obtido no Apêndice 1 em (1.13)

$$\begin{aligned} G''_n(1) &= \mathbb{E}(Z_n^2) - \mathbb{E}(Z_n) \\ &= \text{Var}(Z_n) + \mathbb{E}(Z_n)^2 - \mathbb{E}(Z_n) \end{aligned} \quad \text{de (1.10), no Apêndice 1.}$$

Do Teorema 2.1, resulta que a esperança de Z_n pode ser escrita como μ^n , logo

$$G''_n(1) = \text{Var}(Z_n) + \mu^{2n} - \mu^n. \quad (2.11)$$

2^a. Aplicando a regra da cadeia duas vezes em $G_n(s)$, ou seja, derivando novamente o resultado (2.9)

$$G''_n(s) = G''_{n-1}(G(s))G'(s)G'(s) + G'_{n-1}(G(s))G''(s)$$

e substituindo $s = 1$

$$= G''_{n-1}(1)G'(1)^2 + G'_{n-1}(1)G''(1) \quad \text{de (1.6), } G(1) = 1.$$

A partir da igualdade encontrada no primeiro resultado, (2.11), e sabendo que $G'(1) = \mu$, desenvolve-se o lado direito da equação

$$G_n''(1) = [Var(Z_{n-1}) + \mu^{2(n-1)} - \mu^{n-1}]\mu^2 + G_{n-1}'(1)G''(1).$$

Pelo Teorema 2.1, $G_{n-1}'(1) = \mu^{n-1}$, logo

$$G_n''(1) = [Var(Z_{n-1}) + \mu^{2(n-1)} - \mu^{n-1}]\mu^2 + \mu^{n-1}G''(1).$$

Por fim, o resultado em (2.8) evidenciou que $G''(1)$ pode ser escrito em função de μ e σ^2 , de forma que

$$\begin{aligned} G_n''(1) &= [Var(Z_{n-1}) + \mu^{2(n-1)} - \mu^{n-1}]\mu^2 + \mu^{n-1}(\sigma^2 + \mu^2 - \mu) \\ &= \mu^2 Var(Z_{n-1}) + \mu^{2n} - \mu^{n+1} + \sigma^2 \mu^{n-1} + \mu^{n+1} - \mu^n \\ &= \mu^2 Var(Z_{n-1}) + \mu^{2n} + \sigma^2 \mu^{n-1} - \mu^n. \end{aligned} \quad (2.12)$$

Igualando ambas G_n'' encontradas em (2.11) e (2.12), resulta que a variância de Z_n é igual a

$$\begin{aligned} Var(Z_n) + \mu^{2n} - \mu^n &= \mu^2 Var(Z_{n-1}) + \mu^{2n} + \sigma^2 \mu^{n-1} - \mu^n \\ Var(Z_n) &= \mu^2 Var(Z_{n-1}) + \sigma^2 \mu^{n-1}. \end{aligned}$$

Então, analisando-a quando

- $n = 1$

$$\begin{aligned} Var(Z_1) &= \mu^2 Var(Z_0) + \sigma^2 \mu^0 \\ &= \sigma^2 \end{aligned} \quad \text{por definição, } Var(Z_0) = 0.$$

- $n = 2$

$$\begin{aligned} Var(Z_2) &= \mu^2 Var(Z_1) + \sigma^2 \mu \\ &= \mu^2 \sigma^2 + \sigma^2 \mu \\ &= \sigma^2 \mu^{2-1} (\mu + 1) \times \frac{(\mu - 1)}{(\mu - 1)} \\ &= \sigma^2 \mu^{2-1} \frac{(\mu^2 - 1)}{\mu - 1}. \end{aligned}$$

- $n = 3$

$$\begin{aligned}
\text{Var}(Z_2) &= \mu^2 \text{Var}(Z_2) + \sigma^2 \mu^2 \\
&= \mu^2 \sigma^2 \mu^{3-1} \frac{(\mu^2 - 1)}{\mu - 1} + \sigma^2 \mu^2 \\
&= \sigma^2 \mu^{3-1} \left(\mu \frac{\mu^2 - 1}{\mu - 1} + 1 \right) \\
&= \sigma^2 \mu^{3-1} \left(\frac{\mu^3 - \mu + \mu - 1}{\mu - 1} \right) \\
&= \sigma^2 \mu^{3-1} \left(\frac{\mu^3 - 1}{\mu - 1} \right).
\end{aligned}$$

O resultado será, então, demonstrado por indução, de forma que, supondo válida a generalização para Z_{n-1} , o resultado será válido para Z_n também. Isto é, supõe-se

$$\text{Var}(Z_{n-1}) = \sigma^2 \mu^{(n-1)-1} \left(\frac{\mu^{n-1} - 1}{\mu - 1} \right). \quad (2.13)$$

Dessa forma, resulta que a variância de Z_n é

$$\begin{aligned}
\text{Var}(Z_n) &= \mu^2 \text{Var}(Z_{n-1}) + \sigma^2 \mu^{n-1} \\
&= \mu^2 \sigma^2 \mu^{n-2} \left(\frac{\mu^{n-1} - 1}{\mu - 1} \right) + \sigma^2 \mu^{n-1} && \text{da hipótese (2.13)} \\
&= \sigma^2 \mu^{n-1} \left(\mu \frac{\mu^{n-1} - 1}{\mu - 1} \right) + \sigma^2 \mu^{n-1} \\
&= \sigma^2 \mu^{n-1} \left(\mu \frac{\mu^{n-1} - 1}{\mu - 1} + 1 \right) \\
&= \sigma^2 \mu^{n-1} \left(\frac{\mu^n - \mu + \mu - 1}{\mu - 1} \right) \\
&= \sigma^2 \mu^{n-1} \left(\frac{\mu^n - 1}{\mu - 1} \right).
\end{aligned}$$

□

2.3 Probabilidade de Extinção

Nesta seção será introduzido um conceito importante para o estudo, o de probabilidade de extinção de um processo de ramificação. Mas antes, com o objetivo de sintetizar os resultados vistos até então neste capítulo, considere-se o seguinte exemplo, retirado do livro (GRIMMET; WELSH, 2014).

Seja Z_n um processo de ramificação em que o tamanho das famílias C_j são variáveis

aleatórias independentes e identicamente distribuídas com função de probabilidade

$$p_k = \left(\frac{1}{2}\right)^{k+1}, \text{ para } k = 0, 1, 2, \dots$$

Então, a função geradora de probabilidade de C é

$$G(s) = \sum_{k=0}^{\infty} s^k \left(\frac{1}{2}\right)^{k+1} = \frac{1}{2-s}, \text{ para } |s| < 2.$$

Foi visto na Seção 2.1 que para realizar análises sobre um processo de ramificação, o resultado da equação (2.5) é essencial. Dessa forma, com o objetivo de conhecer o processo em questão, será resolvido, em seguida, $G_n(s) = G_{n-1}(G(s))$.

Primeiramente, procede que se $|s| \leq 1$,

$$G_1(s) = G(s) = \frac{1}{2-s}.$$

Aplicando a equação (2.5) com $n = 2$,

$$\begin{aligned} G_2(s) &= G(G(s)) \\ &= \frac{1}{2 - (2-s)^{-1}} \\ &= \frac{2-s}{3-2s} \quad \text{se } |s| \leq 1. \end{aligned}$$

Para $n = 3$, tem-se

$$\begin{aligned} G_3(s) &= G_2(G(s)) \\ &= \frac{2 - \frac{1}{2-s}}{3 - 2\frac{1}{2-s}} \\ &= \frac{3-2s}{4-3s} \quad \text{se } |s| \leq 1. \end{aligned}$$

De forma geral, escreve-se

$$G_n(s) = \frac{n - (n-1)s}{n + 1 - ns} \quad \text{se } |s| \leq 1.$$

A partir deste resultado é possível, inclusive, definir a função de probabilidade de Z_n . Entretanto, estes cálculos serão omitidos, pois o intuito desta definição para o trabalho é de chegar a resultados a respeito da probabilidade de extinção. Então considere-se a

função de probabilidade de Z_n como

$$\mathbb{P}(Z_n = k) = \begin{cases} \frac{n}{n+1} & \text{se } k = 0, \\ \frac{n^{k+1}}{(n+1)^{k+1}} & \text{se } k = 1, 2, \dots \end{cases}$$

É possível observar que, quando $k = 0$,

$$\mathbb{P}(Z_n = 0) = \frac{n}{n+1} \rightarrow 1, \quad \text{quando } n \rightarrow \infty.$$

Isso significa que, em algum momento, este processo de ramificação será extinto, com probabilidade 1.

Esse exemplo é um caso particular em que o tamanho da família segue uma distribuição geométrica com parâmetro $p = 1/2$. A generalização para qualquer outro parâmetro p , em que $0 < p < 1$, supõe a função de probabilidade de C_j igual a

$$p_k = p(1-p)^k, \quad \text{para } k = 0, 1, 2, \dots$$

Então, obtém-se

$$G(s) = \frac{p}{1 - (1-p)s}, \quad \text{se } |s| < (1-p)^{-1},$$

e a solução da equação (2.5) é

$$G_n(s) = p \frac{(1-p)^n - p^n - (1-p)s[(1-p)^{n-1} - p^{n-1}]}{[(1-p)^{n+1} - p^{n+1}] - (1-p)s[(1-p)^n - p^n]},$$

válida para $n = 1, 2, \dots$ e $|s| \leq 1$. Dessa forma, a probabilidade de extinção ocorre quando

$$\begin{aligned} \mathbb{P}(Z_n = 0) &= G_n(0) && \text{de (1.5)} \\ &= p \frac{(1-p)^n - p^n}{(1-p)^{n+1} - p^{n+1}} \\ &= \frac{\mu^n - 1}{\mu^{n+1} - 1}. \end{aligned}$$

É possível observar que

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_n = 0) = \begin{cases} 1 & \text{se } \mu < 1, \\ \mu^{-1} & \text{se } \mu > 1, \end{cases} \quad (2.14)$$

o que significa que a extinção é garantida caso $\mu < 1$, e não é quando $\mu > 1$. Esse resultado,

junto ao do caso particular visto anteriormente em que $p = 1/2$ e $\mu = (1 - p)/p = 1$, mostram que a extinção de um processo de ramificação cujos tamanhos das famílias seguem a distribuição geométrica, ocorre se e somente se $\mu \leq 1$.

Este não é, contudo, um resultado exclusivo à distribuição do exemplo. É certo que, se o número médio de descendentes deixados por um indivíduo for menor que 1, o processo de ramificação se tornará extinto, qualquer que seja ele (GRIMMET; WELSH, 2014). O Teorema 2.3 abaixo enuncia como encontrar a probabilidade de extinção de um processo de ramificação.

Teorema 2.3. *A probabilidade de extinção \mathbf{e} é a menor raiz não negativa da equação*

$$x = G(x). \quad (2.15)$$

Demonstração. A probabilidade de extinção \mathbf{e} pode ser escrita como

$$\mathbf{e} = \mathbb{P}(Z_n = 0 \text{ para algum } n \geq 0).$$

Seja $E_n = \{Z_n = 0\}$ o evento “o processo de ramificação encontra-se extinto na n -ésima geração”. Então, a probabilidade \mathbf{e}_n de, no momento n , o processo encontrar-se extinto pode ser definida como a probabilidade do evento E_n ,

$$\mathbf{e}_n = \mathbb{P}(E_n).$$

Dessa forma, a extinção do processo de ramificação pode ser escrita como a união dos eventos E_n . Em outras palavras, temos

$$\{Z_n = 0 \text{ para algum } n \geq 0\} = \bigcup_{n=0}^{\infty} E_n.$$

Se $Z_n = 0$, então não existe mais a possibilidade de contaminação em nenhum momento após n , e portanto a geração de infectados seguinte $n + 1$ também possui tamanho 0. Dessa forma, observa-se que $E_n \subseteq E_{n+1}$. Logo, a probabilidade do processo ser extinto na geração n não pode ser maior que a probabilidade em $n + 1$, ou seja, $\mathbf{e}_n \leq \mathbf{e}_{n+1}$.

É possível observar, também, que a sequência de eventos E_n é crescente, ou seja, $E_1 \subseteq E_2 \subseteq \dots$. É adequada, então, a aplicação do Teorema da Continuidade das Medidas de Probabilidade, encontrado em (GRIMMET; WELSH, 2014), para obter

$$\mathbf{e} = \lim_{n \rightarrow \infty} \mathbf{e}_n. \quad (2.16)$$

Como $\mathbf{e}_n = \mathbb{P}(Z_n = 0)$, de 1.5 tem-se que $\mathbf{e}_n = G_n(0)$. De 2.5 e 2.6, sabe-se que

$$\begin{aligned} G_n(s) &= G_{n-1}(G(s)) = G(G(\dots(s)\dots)) \\ &= G(G_{n-1}(s)). \end{aligned}$$

Substituindo $s = 0$, $\mathbf{e}_n = G_n(0)$ satisfaz

$$\mathbf{e}_n = G(\mathbf{e}_{n-1}) \quad \text{para } n = 1, 2, \dots \quad (2.17)$$

sob a condição de que $\mathbf{e}_0 = 0$. Quando $n \rightarrow \infty$, $\mathbf{e}_n \rightarrow \mathbf{e}$. Além disso, G é uma série de potências com raio de convergência de pelo menos 1, uma vez que é contínua no intervalo $[0, 1]$ (GRIMMET; WELSH, 2014). Portanto, é verdade que \mathbf{e} é uma raiz da equação $\mathbf{e} = G(\mathbf{e})$.

Agora, suponha η uma raiz não negativa de (2.15). G é não-decrescente em $[0, 1]$, uma vez que seus coeficientes são não-negativos. Além disso,

$$\begin{aligned} \mathbf{e}_1 &= G(0) \leq G(\eta) = \eta \\ \mathbf{e}_2 &= G(\mathbf{e}_1) \leq G(\eta) = \eta. \end{aligned}$$

Por indução,

$$\mathbf{e}_n \leq \eta \quad \text{para } n = 1, 2, \dots$$

Dessa forma, $\mathbf{e} = \lim_{n \rightarrow \infty} \mathbf{e}_n \leq \eta$, qualquer que seja a raiz de G . \square

O próximo Teorema complementa o resultado desenvolvido acima, evidenciando para quais cenários a extinção é garantida.

Teorema 2.4. *Assuma $p_1 \neq 1$ e $p_0 + p_1 < 1$. A probabilidade \mathbf{e} de extinção satisfaz $\mathbf{e} = 1$ se e somente se a média do tamanho da família μ satisfaz $\mu \leq 1$.*

O caso em que $p_1 = 1$ foi removido porque nele todas as famílias têm tamanho 1, o que satisfaz a condição do teorema e o fere ao mesmo tempo, uma vez que a cadeia não se extingue. Além disso, p_0 será considerado maior que 0 para evitar o caso trivial em que $\mu > 1$ e $\mathbf{e} = 0$.

Demonstração. No intervalo $[0, 1]$, a função geradora de probabilidade G , da variável C ,

- i. é contínua, uma vez que tem raio de convergência maior ou igual que 1,
- ii. não decresce, uma vez que $G'(x) = \sum_k kx^{k-1}p_k \geq 0$, $x \in [0, 1)$,

iii. é convexa, uma vez que $G''(x) = \sum_k k(k-1)x^{k-2}p_k \geq 0$.

Um possível esboço para G é, então, visto na Figura 2. A partir dela, observa-se que, no intervalo $[0, 1]$, a curva $y = G(x)$ e a linha $y = x$ possuem uma ou duas interseções somente. Se a derivada de $G'(1)$ é maior que 1, então a curva de G possui duas interseções com a reta, e $e < 1$, como ilustra o lado esquerdo da figura. Se $G'(1) \leq 1$, a interseção é única, e $x = 1$ é a raiz, para o mesmo intervalo. Contudo, $G'(1) = \mu$ (2.7), então $e = 1$ se e só se $\mu \leq 1$.

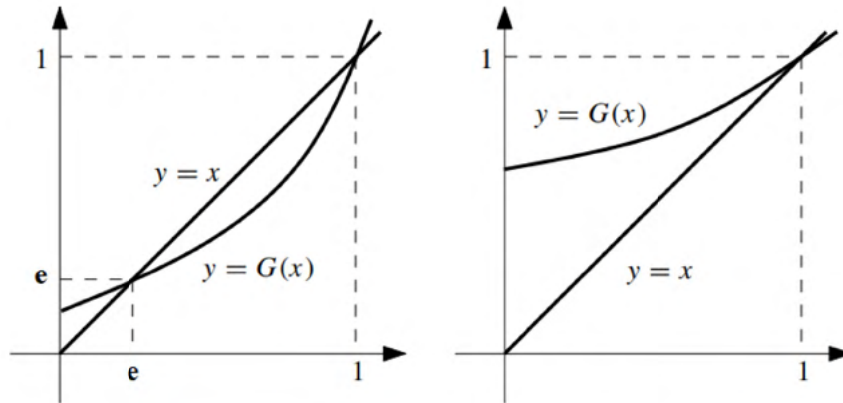


Figura 2: Gráficos da curva $y = G(x)$ com a linha $y = x$ em dois possíveis cenários: à esquerda, com $G'(1) > 1$ e à direita, $G'(1) \leq 1$ (GRIMMET; WELSH, 2014).

□

3 Análise dos Resultados

Neste Capítulo encontram-se os resultados obtidos no estudo de simulação.

Os cálculos e gráficos foram gerados utilizando o software RStudio. Foram feitas 10.000 replicações para cada uma das distribuições de C — Binomial, Geométrica e Poisson. Além disso, para cada distribuição utilizaram-se diferentes definições de parâmetros, afim de estudar o comportamento do processo em situações distintas. É possível acessar os códigos utilizados através do repositório: <https://github.com/g-correia/estudo-de-simulacao-processo-de-ramificacao>.

A Tabela 1 mostra a discriminação dos parâmetros utilizados.

Tabela 1: Resumo das distribuições de C utilizadas na simulação e os respectivos parâmetros definidos.

	Binomial	Geométrica	Poisson
	$m = 10; p = 0,15$	$p = 0,40$	$\lambda = 1,5$
	$m = 10; p = 0,03$	$p = 0,65$	$\lambda = 0,3$
	$m = 10; p = 0,10$	$p = 0,50$	$\lambda = 1,0$

Os resultados encontrados são apresentados nas seções deste capítulo, separadas por distribuição, onde se apresentam análises a respeito de média, variância e probabilidade de extinção.

3.1 Binomial

Nesta seção serão analisados os resultados quando a distribuição de C é Binomial.

A Figura 3 mostra, para cada instante n da epidemia, a evolução das médias teórica e estimada pela simulação do número de infectados. Em todos os casos, é possível notar uma concordância entre valor teórico e observado. No gráfico do item (a), tem-se $\mu = 1,5$, maior que um. Neste cenário, o comportamento do gráfico é crescente para ambas as medidas, e a quantidade média de infectados encontra-se na casa dos milhares entre as

gerações 15 e 20. Em (b), quando $\mu = 0,3$, é possível observar um decrescimento rápido, com extinção a partir da geração 5. No gráfico superior do item (c), não se observa a extinção nem o crescimento do número médio de infectados, ele se distancia muito pouco da média; este é o caso em que $\mu = 1$. No gráfico inferior, por sua vez, a simulação é repetida para a geração $n = 100$, onde a média fica abaixo de 1.

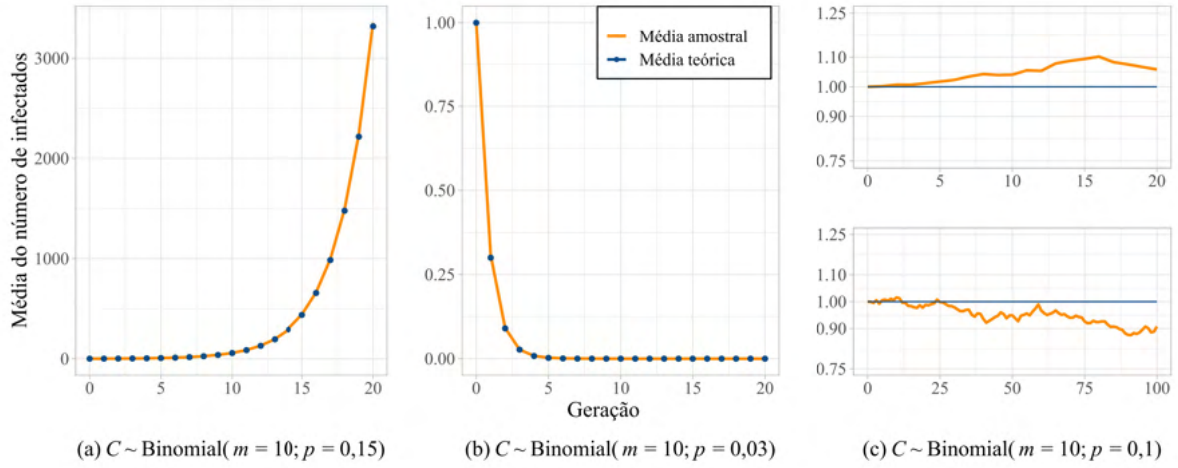


Figura 3: Gráficos das médias teórica e amostral para C com distribuição Binomial de diferentes parâmetros ao longo das gerações.

A Figura 4 mostra a comparação dos resultados teórico e observado da variância do número de infectados ao longo das gerações. Assim como ocorreu com a média, os valores obtidos também estão de acordo com a literatura. Para μ maior que um, a variabilidade tende a aumentar conforme n aumenta. O inverso ocorre quando $\mu < 1$, como mostra o gráfico do item (b); a variabilidade é grande nas primeiras gerações, e depois tende a zero, conforme a cadeia também tende à extinção. Para $\mu = 1$, é possível ver no item (c) que a variabilidade do número de infectados aumenta linearmente, inclusive para n grande.

Para analisar a probabilidade de extinção quando $\mu = 1,5$ no item (a) da Figura 5, primeiro calculou-se o valor máximo que a probabilidade pode atingir a partir do Teorema 2.3, de onde foi obtida a função

$$\begin{aligned} x &= G(x) \\ x &= (0,85 + 1,5x)^{10} && \text{de (1.7), no Apêndice 1} \\ x - (0,85 + 1,5x)^{10} &= 0. \end{aligned}$$

A menor raiz da função foi, então, encontrada a partir do método numérico utilizado para aproximação de raízes de funções reais contínuas denominado Método da Bissecção, obtendo-se $e \simeq 0,3715$. Neste caso não há garantia da epidemia se extinguir, uma vez

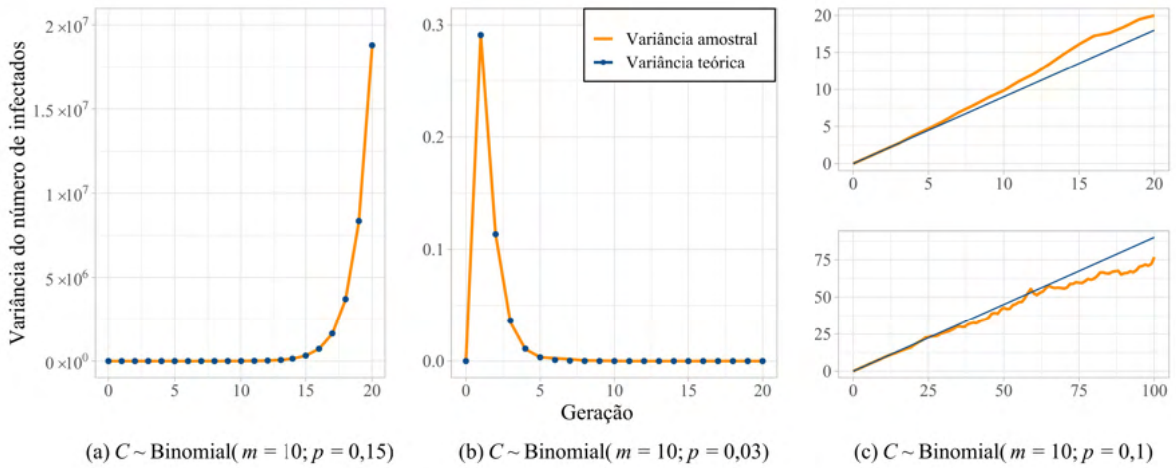


Figura 4: Gráficos das variâncias teórica e amostral para C com distribuição Binomial de diferentes parâmetros ao longo das gerações.

que a probabilidade de extinção converge para o valor calculado, menor que um. Nos itens (b) e (c), é possível observar que a probabilidade de extinção converge para um, assim como apontado na literatura. O último item mostra, também, a necessidade de um tempo maior para que a probabilidade de extinção convirja para 1, em oposição ao item (b), para $\mu < 1$, quando esta probabilidade converge mais rapidamente. Isto significa que a probabilidade de extinção é garantida mais cedo quando $\mu < 1$.

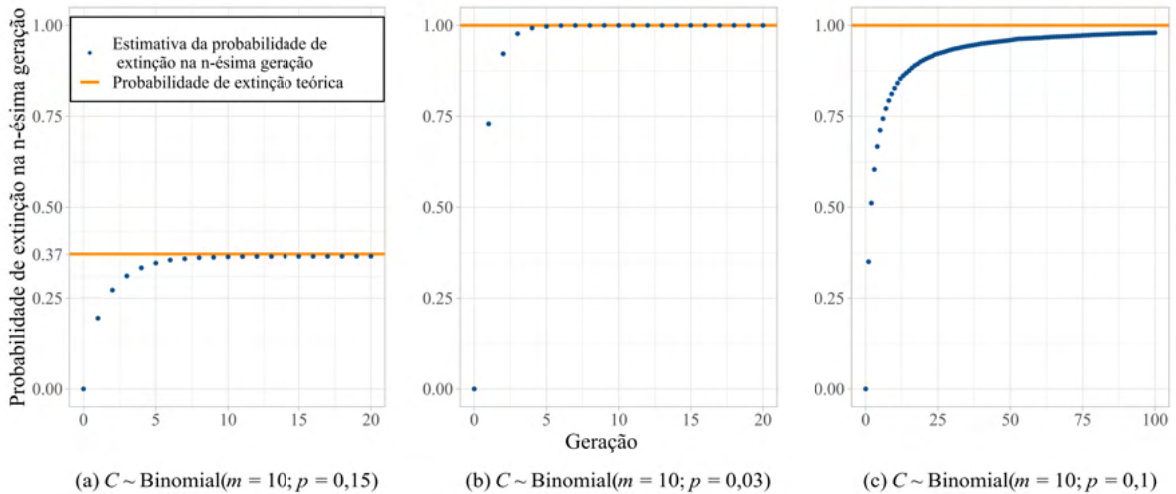


Figura 5: Gráficos das probabilidades de extinção na n -ésima geração teórica e amostral para C com distribuição Binomial de diferentes parâmetros.

3.2 Geométrica

Nesta seção será desenvolvido o comportamento do processo quando C tem distribuição Geométrica.

A Figura 6 mostra que a evolução das médias observadas com a simulação também estão próximas do valor teórico para o caso da distribuição Geométrica. No gráfico do item (a) tem-se a mesma média $\mu = 1,5$ que o item (a) da Figura 3, e além do crescimento, é possível observar que os gráficos apresentam uma média de infectados muito semelhantes em todas as gerações; o crescimento se torna aparente pouco antes da 10^a geração, alcança os 500 infectados na 15^a e finaliza na 20^a geração em torno de 3000 infectados. Ou seja, para ambas distribuições Binomial e Geométrica, a progressão da epidemia é muito semelhante quando o número médio de infectados é o mesmo. Com o item (b), quando $\mu \simeq 0,54$, conclui-se que a epidemia apresenta um ritmo decrescente desde as primeiras gerações. Uma outra característica presente neste gráfico é que, com relação ao gráfico de mesmo item da Figura 3, este apresenta um decrescimento menos acelerado. Isto se deve ao fato da média de infectados ser um pouco maior que no caso anterior. Nos gráficos do item (c), onde $\mu = 1$, mais uma vez não se observa o crescimento do número médio de infectados nem a extinção, mas um comportamento mais constante próximo da média. O gráfico inferior evidencia um distanciamento maior entre o valor amostral e o valor esperado para gerações mais avançadas, quando comparado com as gerações iniciais no gráfico superior. Ainda assim, a oscilação fica dentro de um intervalo pequeno, de menos que 0,25.

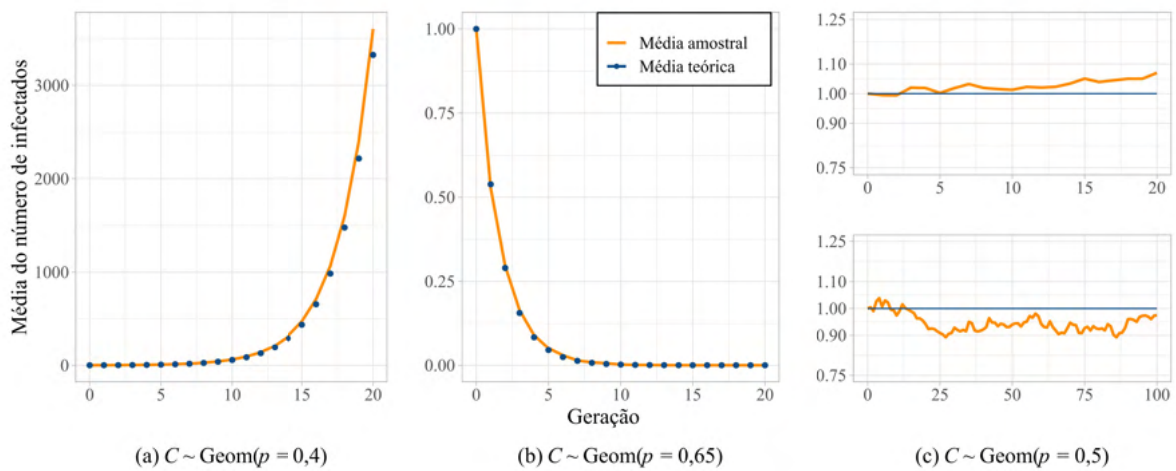


Figura 6: Gráficos das médias teórica e amostral para C com distribuição Geométrica de diferentes parâmetros ao longo das gerações.

Para as análises de variância na Figura 7, assim como no caso da distribuição Binomial visto na Seção 3.1, observa-se no item (a) que para uma epidemia em estado de crescimento, com $\mu > 1$, a variabilidade aumenta à proporção que o número de infectados alcança valores maiores. No caso inverso, apresentado no item (b), a variabilidade também é proeminente nas primeiras gerações quando $n \neq 0$. A maior diferença entre os

casos da Geométrica e da Binomial está na comparação entre itens (a), os quais apresentam mesma média do número de infectados, $\mu = 1,5$. Nele, mesmo com uma evolução do número de infectados similar, a distribuição Geométrica apresenta variabilidade superior devido a natureza da distribuição.

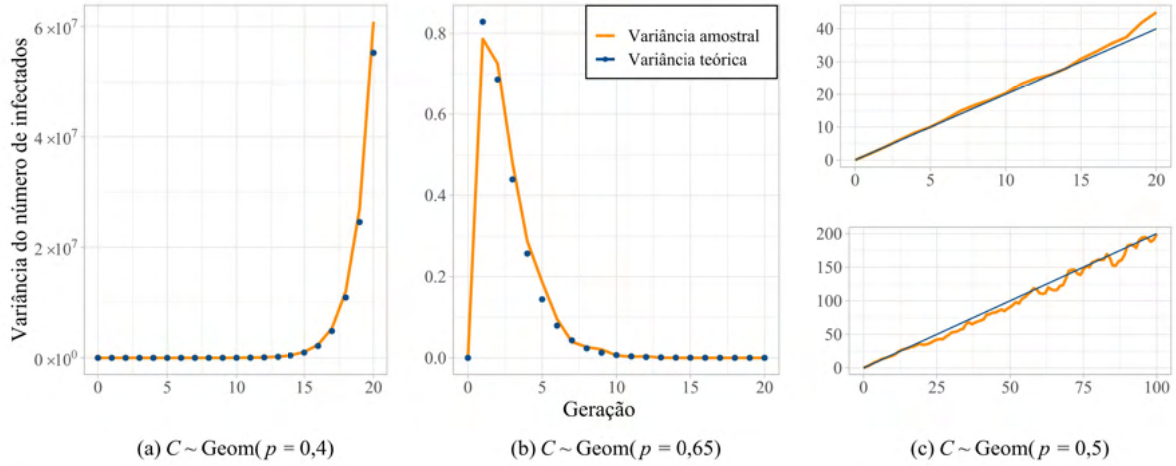


Figura 7: Gráficos das variâncias teórica e amostral para C com distribuição Geométrica de diferentes parâmetros ao longo das gerações.

A Figura 8 mostra no item (a) que para $\mu > 1$ não há garantia da epidemia se extinguir, uma vez que a probabilidade de extinção converge para o valor calculado no resultado (2.14), $\mu^{-1} \simeq 0,6667$. Nos itens (b) e (c), a probabilidade de extinção converge para um, como era esperado. Mas o último item evidencia que quando $\mu = 1$, é necessário mais tempo para que o valor se aproxime de 1, em oposição ao item (b), que mostra uma convergência mais rápida quando $\mu < 1$. Adicionalmente, quando $\mu > 1$, a distribuição Binomial apresentou menor probabilidade de extinção ($e \simeq 0,3715$) que a distribuição Geométrica ($e \simeq 0,6667$), apesar de mesma média do número de infectados, $\mu = 1,5$.

3.3 Poisson

Para a última distribuição analisada, a Figura 9 mostra que a evolução das médias observadas na simulação também estão próximas do valor teórico. Mais uma vez, o item (a) apresenta um padrão de crescimento muito semelhante aos demais itens (a) mostrados anteriormente neste Capítulo, nos quais a média é $\mu = 1,5$. Ou seja, para todas as distribuições analisadas no estudo, a progressão do número médio de infectados na epidemia é muito semelhante quando μ é o mesmo. Com o item (b), quando $\mu = 0,3$, conclui-se que a epidemia apresenta um ritmo decrescente desde as primeiras gerações, com extinção em torno da quinta geração, tal qual a distribuição Binomial de mesmo μ .

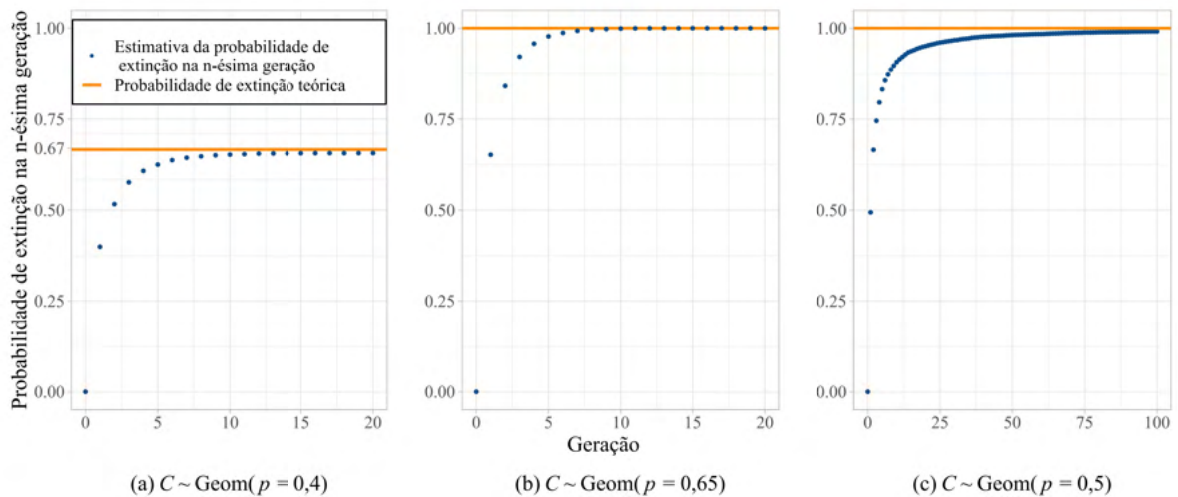


Figura 8: Gráficos das probabilidades de extinção na n -ésima geração teórica e amostral para C com distribuição Geométrica de diferentes parâmetros.

Nos gráficos do item (c), onde $\mu = 1$, está presente o comportamento constante em torno da média, com o número de infectados se mantendo dentro do intervalo $[0, 9; 1, 1]$.

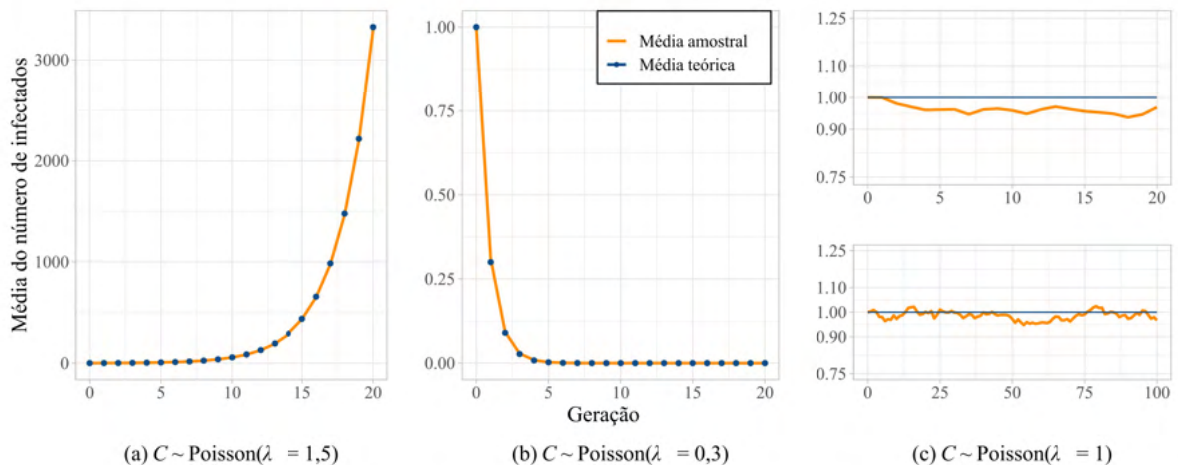


Figura 9: Gráficos das médias teórica e amostral para C com distribuição Poisson de diferentes parâmetros ao longo das gerações.

O item (a) da Figura 10 mostra que para uma epidemia em estado de crescimento, com $\mu > 1$, a variabilidade aumenta exponencialmente à medida que o número de infectados alcança valores maiores. No item (b), quando a média é menor que 1, a variabilidade se mostra proeminente somente nas primeiras gerações, e depois tende a zero quando a epidemia também tende a extinção. No item (c), a variabilidade aumenta linearmente tal qual o valor teórico, inclusive para n grande. De forma geral, a variância se comporta de forma semelhante ao caso da distribuição Binomial visto na Seção 3.1.

O valor de convergência da probabilidade de extinção quando $\mu = 1, 5$ no item (a) da

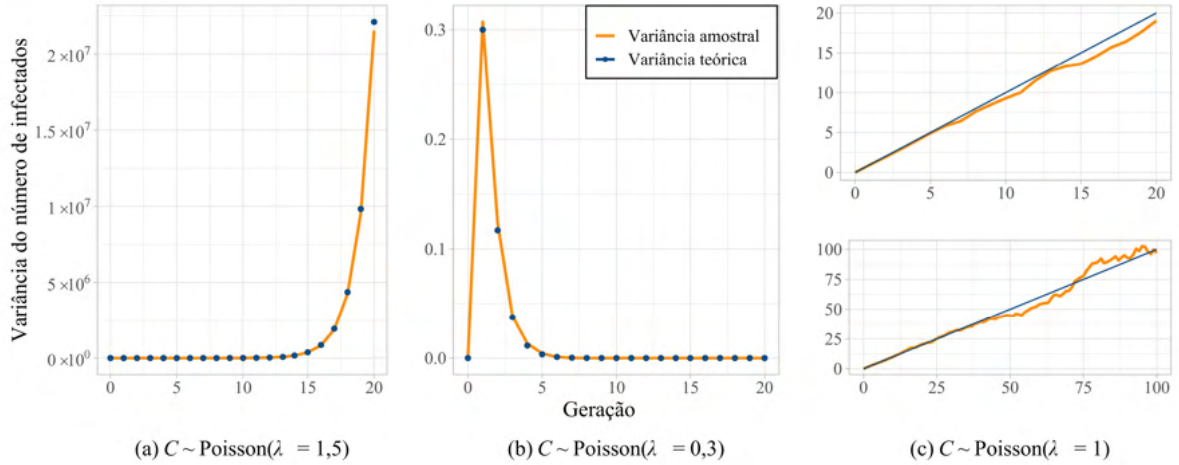


Figura 10: Gráficos das variâncias teórica e amostral para C com distribuição Poisson de diferentes parâmetros ao longo das gerações.

Figura 11 foi calculado da mesma forma que na seção 3.1, a partir do Teorema 2.3, de onde obteve-se

$$\begin{aligned}
 x &= G(x) \\
 x &= e^{1,5(x-1)} && \text{de (1.8), no Apêndice 1} \\
 x - e^{1,5(x-1)} &= 0.
 \end{aligned}$$

A menor raiz da função foi encontrada através do Método da Bissecção, obtendo-se $e \simeq 0,41718$. Neste caso, não há garantia da epidemia se extinguir, uma vez que a probabilidade de extinção é, no máximo, o valor calculado. Nos itens (b) e (c), é possível observar que a probabilidade de extinção converge para um, assim como mostra a literatura. O último item evidencia, também, a necessidade de um tempo maior para que a probabilidade de extinção convirja para 1, em oposição ao item (b), para $\mu < 1$, quando esta probabilidade converge mais rapidamente. Isto significa que a probabilidade de extinção é garantida mais cedo quando $\mu < 1$ em oposição a $\mu = 1$.

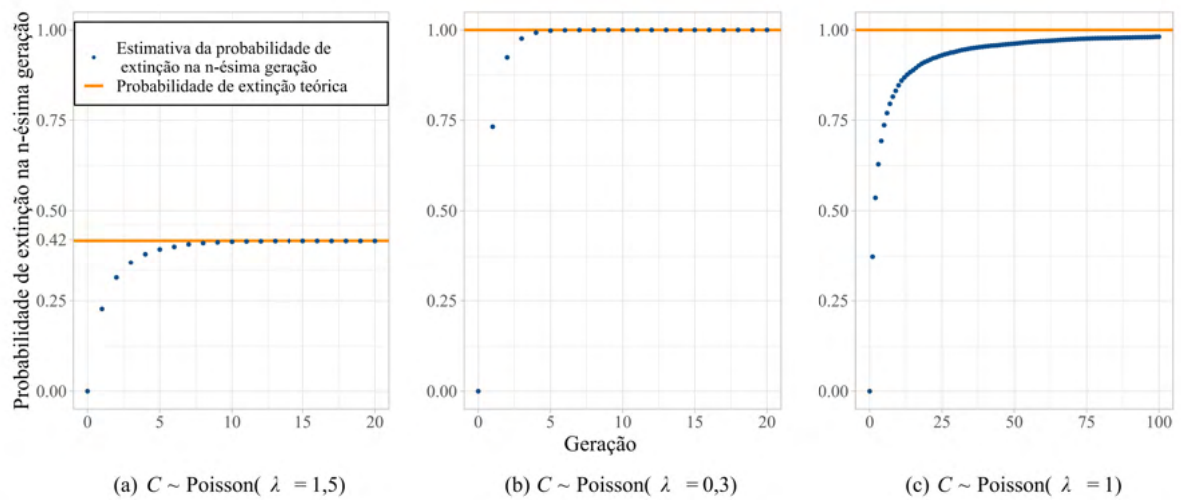


Figura 11: Gráficos das probabilidades de extinção na n -ésima geração teórica e amostral para C com distribuição Poisson de diferentes parâmetros.

4 Conclusão

Diante dos impactos emergenciais que o surgimento de epidemias como a da COVID-19 podem acender, marcados pela carência de informações a respeito da propagação da doença nova, surge a motivação para o estudo de um método eficaz e de simples provimento. Desta forma, o objetivo geral deste trabalho de conclusão de curso foi estudar a teoria do processo de ramificação com aplicações voltadas para epidemiologia, bem como o desenvolvimento de uma aplicação prática, exercida por meio de um estudo de simulação. Para isso, foi também explorado o desempenho de diferentes distribuições — Binomial, Geométrica e Poisson — para a estipulação do número de infectados por um indivíduo contaminado, representado pela variável aleatória C .

Nos resultados obtidos com a simulação, mostraram-se análises a respeito do número médio de infectados deixados por um único indivíduo, μ , e probabilidade de extinção da epidemia, e , ao longo das gerações, n . Para $\mu < 1$, a média do número total de infectados a cada geração tende a decrescer, e a probabilidade de extinção é crescente, tendendo a ser garantida passado o tempo. Em contrapartida, para $\mu > 1$, foi observado que a média aumenta a cada geração, e que a probabilidade de extinção existe e é mensurável, porém não garantida. Paralelamente, para $\mu = 1$, os resultados apontam para a estabilidade do número total de infectados em torno de um, e para uma probabilidade de extinção que tende a se certificar quanto mais gerações se atingem. Estes resultados denotam uma concordância marcante entre teoria e prática. Uma outra conclusão observada no estudo foi de que a natureza da distribuição que se escolhe para C impacta nos resultados, em especial, na variância do número médio de infectados por geração.

Entre as considerações pessoais derivadas do trabalho, destaca-se o aproveitamento proporcionado pela investigação de uma técnica de modelagem estatística diferente das contempladas ao longo da graduação, que intensificou a indagação a respeito das possibilidades de estudo proporcionadas pelo curso. Deparou-se mais uma vez com a conveniência do aprofundamento no estudo das distribuições de probabilidade, o que permanece como um dos pontos a ser aperfeiçoado após esta etapa. Fez-se saudososa a diretriz de base

epidemiológica, que poderia ter sido um fator de relevância para o trabalho. Contudo, os aspectos teóricos aprofundados, em especial a força do resultado da probabilidade de extinção e a diferença de comportamento entre cada cenário abordado, foram grandes novidades aproveitadas durante este trabalho.

Como indicação de trabalhos futuros, acredita-se que o estudo do modelo de ramificação sob abordagem de cunho ainda mais epidemiológico poderia trazer resultados de vigilância epidemiológica relevantes. Neste sentido, poderiam-se explorar a alimentação do modelo utilizando dados reais sobre a COVID-19, referências de ajuste de parâmetro próprias da área de acordo com tipos de doenças diferentes, bem como análises do desempenho desse tipo de modelo em aplicações preditivas.

Referências

- ANDERSSON, H.; BRITTON, T. *Stochastic Epidemic Models and Their Statistical Analysis*. 1. ed. [S.l.]: Springer New York, NY, 2000.
- BORGES, R. L. e. R. S. B. Análise estocástica de propagação de doenças epidemiológicas. *Revista Mundi Engenharia, Tecnologia e Gestão*, v. 6, n. 3, 2021.
- FARRINGTON, C.; KANAAN, M.; GAY, N. Branching process models for surveillance of infectious diseases controlled by mass vaccination. *Biostatistics (Oxford, England)*, v. 4, p. 279–95, 05 2003.
- FRANCO, C. M. R.; DUTRA, R. F. Sir model for propagation of covid-19 in the paraíba's state (brazil). *INTERMATHS*, v. 2, n. 2, p. 39–48, Dec. 2021.
- GRIMMET, G.; WELSH, D. *Probability An Introduction*. 2. ed. [S.l.]: Oxford University Press, 2014.
- MOHAPATRA, R. K. et al. The recent challenges of highly contagious covid-19, causing respiratory infections: Symptoms, diagnosis, transmission, possible vaccines, animal models, and immunotherapy. *Chemical Biology & Drug Design*, v. 96, n. 5, p. 1187–1208, 2020.
- ORGANIZATION, W. H. *Transmission of SARS-CoV-2: implications for infection prevention precautions: scientific brief, 09 July 2020*. [S.l.], 2020. 10 p. p.
- ZHAO, T. et al. The importance of social distancing policy. *Journal of public health*, v. 43 2, p. e269, 2021.

APÊNDICE 1

Neste apêndice estão apresentados resultados e definições importantes utilizados ao longo do trabalho.

1.1 Momentos de uma variável aleatória discreta

Para todos os resultados mostrados nessa seção, X será considerada uma variável aleatória discreta, com média μ e variância σ^2 .

Definição 1.1. *A esperança de X é denotada por $\mathbb{E}(X)$, e é dada por*

$$\mathbb{E}(X) = \sum_{x \in \text{Im}X} x\mathbb{P}(X = x).$$

Dessa definição, é possível perceber que $\mathbb{E}(X)$ é a média de X . Em outras palavras,

$$E(X) = \mu. \tag{1.1}$$

Definição 1.2. *A variância de X denotada por $\text{Var}(X)$ é dada por*

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]. \tag{1.2}$$

Da Definição 1.2, tem-se a seguinte proposição

Proposição 1.1.

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2. \tag{1.3}$$

Demonstração.

$$\begin{aligned}
 \text{Var}(X) &= \mathbb{E}[(X - \mathbb{E}(X))^2] \\
 &= \mathbb{E}[X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2] \\
 &= \mathbb{E}(X^2) - \mathbb{E}[2X\mathbb{E}(X)] + \mathbb{E}(X)^2 \\
 &= \mathbb{E}(X^2) - \sum_x 2x\mathbb{E}(X)p_x(x) + \mathbb{E}(X)^2 \\
 &= \mathbb{E}(X^2) - 2\mathbb{E}(X) \sum_x xp_x(x) + \mathbb{E}(X)^2 \\
 &= \mathbb{E}(X^2) - 2\mathbb{E}(X)^2 + \mathbb{E}(X)^2 \\
 &= \mathbb{E}(X^2) - \mathbb{E}(X)^2.
 \end{aligned}$$

□

1.2 Função Geradora de Probabilidade

1.2.1 Definição

Definição 1.3. *Seja X uma variável aleatória discreta, onde $P(X = x) = p_x$. A função geradora de probabilidade de X é definida como*

$$\mathbb{E}(s^X) = G_X(s) = p_0 + p_1s + p_2s^2 + \dots \quad (1.4)$$

para todos os valores de s que a sequência converge absolutamente.

Abaixo estão casos particulares desta função, que ocorrem quando $s = 0$ e $s = 1$

$$G_X(0) = p_0, \quad (1.5)$$

$$G_X(1) = p_0 + p_1 + p_2 + \dots = 1. \quad (1.6)$$

A seguir são apresentados exemplos das funções geradoras de probabilidade para uma variável aleatória X , caso esta assuma distribuição Binomial ou Poisson.

Exemplo 1. Binomial com parâmetros m e p

Então

$$P(X = k) = \binom{m}{k} p^k (1-p)^{m-k}, \quad k = 0, 1, \dots, m.$$

Logo, a função geradora de probabilidade de X é definida como

$$G_X(s) = \sum_{k=0}^m \binom{m}{k} p^k (1-p)^{m-k} s^k = (1-p+ps)^m. \quad (1.7)$$

Exemplo 2. Poisson com parâmetro λ .

Então

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, \dots$$

Logo, a função geradora de probabilidade de X é definida como

$$G_X(s) = \sum_{k=0}^{\infty} \frac{1}{k!} \lambda^k e^{-\lambda} s^k = e^{\lambda(s-1)}. \quad (1.8)$$

1.2.2 Momentos

Definição 1.4. Seja $k \geq 1$. O k -ésimo momento de X é a quantidade $\mathbb{E}(X^k)$.

Do resultado acima, é possível escrever a média e a variância de X em função do 1º e 2º momentos, de forma que μ é o primeiro momento, e σ^2 , a diferença entre o segundo momento e o quadrado do primeiro

$$\mu = \mathbb{E}(X), \quad (1.9)$$

$$\sigma^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2. \quad (1.10)$$

Teorema 1.1. Seja X uma variável aleatória com função geradora de probabilidade $G_X(s)$. A r -ésima derivada de $G_X(s)$ quando $s = 1$ é $\mathbb{E}[X \cdot (X - 1) \cdot \dots \cdot (X - r + 1)]$, para $r = 1, 2, \dots$. Ou seja,

$$G_X^{(r)}(1) = \mathbb{E}[X \cdot (X - 1) \cdot \dots \cdot (X - r + 1)].$$

Demonstração. Para $r = 1$,

$$\begin{aligned} G_X'(s) &= \frac{d}{ds} \sum_{k=0}^{\infty} s^k \mathbb{P}(X = k) \\ &= \sum_{k=0}^{\infty} \frac{d}{ds} s^k \mathbb{P}(X = k) \\ &= \sum_{k=0}^{\infty} k s^{k-1} \mathbb{P}(X = k). \end{aligned}$$

Então, para $s = 1$

$$G'_X(1) = \sum_{k=0}^{\infty} k\mathbb{P}(X = k) = \mathbb{E}(X).$$

Para a segunda derivada,

$$\begin{aligned} G''_X(s) &= \frac{d}{ds} \sum_{k=0}^{\infty} k s^{k-1} \mathbb{P}(X = k) \\ &= \sum_{k=0}^{\infty} \frac{d}{ds} k s^{k-1} \mathbb{P}(X = k) \\ &= \sum_{k=0}^{\infty} k(k-1) s^{k-2} \mathbb{P}(X = k) \end{aligned}$$

que substituindo $s = 1$ resulta em

$$G''_X(1) = \sum_{k=0}^{\infty} k(k-1) \mathbb{P}(X = k) = \mathbb{E}[X(X-1)].$$

Por indução, então, supõe-se válido que a derivada de índice $r-1$ da função geradora de probabilidade de X seja igual a

$$G_X^{(r-1)}(s) = \sum_{k=0}^{\infty} k(k-1)\dots(k-r) s^{k-r-1} \mathbb{P}(X = k).$$

Dessa forma, a r -ésima derivada de G_X pode ser escrita como

$$\begin{aligned} G_X^{(r)}(s) &= \frac{d}{ds} \sum_{k=0}^{\infty} G_X^{(r-1)}(s) \\ &= \sum_{k=0}^{\infty} k(k-1)\dots(k-r)(k-r-1) s^{k-r-2} \mathbb{P}(X = k). \end{aligned}$$

E quando $s = 1$,

$$\begin{aligned} G_X^{(r)}(1) &= \sum_{k=0}^{\infty} k(k-1)\dots(k-r)(k-r-1) \mathbb{P}(X = k) \\ &= \mathbb{E}[X \cdot (X-1) \cdot \dots \cdot (X-r-1)]. \end{aligned}$$

□

Dessa definição, podemos calcular o primeiro momento de X

$$\mathbb{E}(X) = G'_X(1) \tag{1.11}$$

e o segundo momento de X

$$\begin{aligned}\mathbb{E}(X^2) &= \mathbb{E}[X(X-1) + X] \\ &= \mathbb{E}[X(X-1)] + \mathbb{E}(X) \\ &= G_X''(1) + G_X'(1).\end{aligned}\tag{1.12}$$

Unindo os dois últimos resultados obtidos, então, define-se a segunda derivada de $G_X(1)$

$$G_X''(1) = \mathbb{E}(X^2) - \mathbb{E}(X).\tag{1.13}$$

Teorema 1.2. *Se X e Y são variáveis aleatórias independentes, ambas com imagem em $\{0, 1, 2, \dots\}$, então a soma delas tem função geradora de probabilidade*

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

Demonstração.

$$\begin{aligned}G_{X+Y}(s) &= \mathbb{E}(s^{X+Y}) = \mathbb{E}(s^X s^Y) \\ &= \mathbb{E}(s^X)\mathbb{E}(s^Y) \\ &= G_X(s)G_Y(s).\end{aligned}$$

□

Esse resultado pode ser generalizado para a soma $S_n = X_1 + X_2 + \dots + X_n$ de n variáveis aleatórias independentes, todas com imagem em $\{0, 1, 2, \dots\}$, de forma que

$$G_{S_n}(s) = G_{X_1}(s)G_{X_2}(s)\dots G_{X_n}(s).\tag{1.14}$$

Teorema 1.3. *Sejam N, X_1, X_2, \dots variáveis aleatórias independentes, todas com imagem em $\{0, 1, 2, \dots\}$. Se X_i são identicamente distribuídas, com mesma função geradora de probabilidade G_X , então a soma*

$$S = X_1 + X_2 + \dots + X_N$$

tem função geradora de probabilidade

$$G_S(s) = G_N(G_X(s)).\tag{1.15}$$

Demonstração. Para provar este resultado, será necessário o auxílio do Teorema da Partição, enunciado abaixo.

Teorema 1.4. *Se X é uma variável discreta, e $\{B_1, B_2, \dots\}$ é uma partição do espaço amostral, tal que $\mathbb{P}(B_i) > 0$ para cada i , então*

$$\mathbb{E}(X) = \sum_i \mathbb{E}(X|B_i)\mathbb{P}(B_i),$$

sempre que essa soma converge absolutamente.

Então, a partir dos eventos $B_n = \{N = n\}$, é possível concluir que

$$\begin{aligned} G_S(s) &= \mathbb{E}(s^{X_1+\dots+X_N}) \\ &= \sum_{n=0}^{\infty} \mathbb{E}(s^{X_1+\dots+X_N} | N = n) \mathbb{P}(N = n) && \text{do Teorema 1.4} \\ &= \sum_{n=0}^{\infty} \mathbb{E}(s^{X_1+\dots+X_N}) \mathbb{P}(N = n) \\ &= \sum_{n=0}^{\infty} G_X(s)^n \mathbb{P}(N = n) && \text{de (1.14)} \\ &= G_N(G_X(s)). \end{aligned}$$

□