

Modelos de crescimento exponencial, logístico e Gompertz: Uma revisão teórica e aplicação a dados da COVID-19

A. S. GUIMARÃES¹, M. I. C. M. ROSSI², W. A. O. SOLER³ e E. F. SARAIVA^{4*}

Recebido em 17 de outubro de 2024 / Aceito em 23 de março de 2025

RESUMO. O cenário pandêmico causado pelo coronavírus SARS-CoV-2 fez aumentar o interesse e a divulgação de modelos matemáticos capazes de projetar a evolução do número de casos (e mortes) devido à COVID-19, em países, estados e cidades. Em muitos artigos, o número acumulado de casos é modelado por um modelo de crescimento não-linear, tais como os modelos exponencial, logístico e de Gompertz. Motivados por este fato, neste artigo, apresentamos uma revisão detalhada sobre estes três modelos de crescimento. Iniciamos com a obtenção da expressão matemática do modelo exponencial utilizando um exemplo simples de divisão celular. A partir do modelo exponencial, apresentamos em detalhes o desenvolvimento matemático para obtenção das expressões dos modelos logístico e de Gompertz, e como obter as coordenadas do ponto de inflexão deste dois modelos. Apresentada a revisão, ilustramos o uso destes três modelos na modelagem do número acumulado de mortes por COVID-19 registradas no estado de São Paulo no período de 17/03/2020 a 30/04/2021. Nesta modelagem, apresentamos um critério para o ajuste de um modelo por partes haja vista que os valores registrados apresentam um comportamento heterogêneo ao longo do tempo.

Palavras-chave: modelo exponencial, modelo logístico, modelo de Gompertz, ponto de inflexão, modelagem estatística, estimação.

1 INTRODUÇÃO

No mês de Dezembro do ano de 2019, um novo coronavírus, denominado de SARS-Cov-2, foi descoberto na cidade de Wuhan, China [13]. A partir da sua descoberta, o vírus se espalhou rapidamente pelo mundo e se tornou um dos principais problemas de saúde, causando um

*Autor correspondente: Erlandson F. Saraiva – E-mail: erlandson.saraiva@ufms.br

¹Instituto de Matemática, Universidade Federal de Mato Grosso do Sul, 79090-900, Campo Grande, MS, Brasil – E-mail: abraao.guimaraes@ufms.br <https://orcid.org/0009-0006-3163-9925>

²Instituto de Matemática, Universidade Federal de Mato Grosso do Sul, 79090-900, Campo Grande, MS, Brasil – E-mail: maria.rossi@ufms.br <https://orcid.org/0000-0002-5415-5572>

³Instituto de Matemática, Universidade Federal de Mato Grosso do Sul, 79090-900, Campo Grande, MS, Brasil – E-mail: willy.oliveira@ufms.br <https://orcid.org/0000-0002-7467-2307>

⁴Instituto de Matemática, Universidade Federal de Mato Grosso do Sul, 79090-900, Campo Grande, MS, Brasil – E-mail: erlandson.saraiva@ufms.br <https://orcid.org/0000-0002-4698-4004>

enorme ônus social e econômico. Em 11 de Março de 2020 a COVID-19 foi declarada como uma pandemia pela organização mundial da saúde [8]. Somente a partir do desenvolvimento e da disponibilização em massa das vacinas contra este coronavírus é que foi possível reduzir o número de contágios e mortes.

No período da pandemia, houve um aumento no interesse e na divulgação de modelos matemáticos capazes de projetar a evolução do número de casos (e mortes) devido à COVID-19 em países, estados e cidades. Esse interesse se deve principalmente ao fato de que as projeções podem auxiliar os agentes governamentais na tomada de decisões, tais como, intensificação do isolamento social, aquisição de equipamentos hospitalares, aumento do número de unidades de terapia intensiva nos hospitais, aplicação de um *lockdown*, entre outras.

Em geral, os artigos publicados, modelam o número acumulado de casos ou mortes, pois estas medidas apresentam um comportamento mais estável em relação às quantidades registradas por dia [12]. Sob este cenário, três modelos de crescimento muito utilizados foram: o modelo exponencial, o modelo logístico [2, 14] e o modelo de Gompertz [5]. Por exemplo, [11] utiliza o modelo exponencial para modelar o número de casos registrados na Itália no período de 21 de Fevereiro de 2020 a 8 de março de 2020. Em [9], o modelo exponencial é considerado para analisar o número de casos registrados na África no período de 1^o de Março à 13 de Abril de 2020. Em [6], são utilizados os modelos exponencial e logístico para modelar o número de casos registrados em diversos estados da Índia no período de 30 de Janeiro de 2020 à 3 de Maio de 2020. Em [8], são utilizados os modelos exponencial e de Gompertz para modelar o número de casos registrados na Índia, Paquistão, Mianmar, Brasil, Itália e Alemanha no período de 30 de Janeiro de 2020 à de junho de 2020.

Uma particularidade dos artigos que utilizaram os modelos exponencial, logístico e de Gompertz para modelar os dados da COVID-19, é que os autores apenas assumem um dos modelos sem apresentar alguma discussão sobre o por quê de aplicá-los. O que é natural, pois são modelos bem conhecidos da literatura. Contudo, não há disponível na literatura uma revisão detalhada destes três modelos em um único artigo, apresentando a construção do modelo exponencial e como os modelos logístico e de Gompertz são obtidos a partir do modelo exponencial, apenas inserindo um mecanismo limitador do crescimento populacional.

Motivados por este fato, neste artigo, apresentamos de maneira didática e detalhada a construção destes três modelos de crescimento. Para isto, apresentamos os modelos em sua forma de equação diferencial e descrevemos em detalhes a solução destas equações para obter a expressão matemática conhecida de cada um dos modelos. Para os modelos logístico e de Gompertz, também descrevemos os procedimentos matemáticos para obtenção das coordenadas do ponto de inflexão. O interesse no ponto de inflexão se deve ao fato de que este ponto é o ponto da curva em que a taxa de crescimento atinge o valor máximo.

Descrito os três modelos, ilustramos o seu uso na modelagem do número acumulado de mortes por COVID-19 registrados no estado de São Paulo no período de 17/03/2020 a 30/04/2021. Este dados foram obtidos de forma gratuita no website <http://www.seade.gov.br/coronavirus>.

Para isto, assumimos que o logaritmo das quantidades acumuladas foram geradas de acordo com um modelo aditivo composto por uma função de crescimento (log-exponencial, log-logística ou log-Gompertz), representando a parte esperada do modelo aditivo, mais de um erro aleatório assumido como sendo gerado de uma distribuição normal com média 0 e variância σ^2 . A opção pela modelagem do logaritmo das quantidades acumuladas se deve ao fato de reduzir a escala dos valores registrados, evitando assim problemas computacionais no processo de estimação dos parâmetros, e também para podermos considerar a suposição de normalidade para os erros aleatórios, uma vez que as medidas log-transformadas passam a ser medidas contínuas em \mathbb{R} . Para obter as estimativas para os parâmetros dos modelos, consideramos o método dos mínimos quadrados não-linear, conforme descrito por [15].

Ajustamos os três modelos aos dados referentes ao logaritmo do número acumulado de mortes registrado do período de 17/03/2020 a 30/04/2021 no estado de São Paulo e selecionamos o melhor modelo utilizando os critérios de seleção de modelos AIC e BIC e o erro quadrático médio (EQM). Estes três critérios indicaram o modelo de Gompertz como sendo o melhor modelo, *i.e.*, o modelo que melhor explica os dados observados dentre os modelos testados.

No entanto, como o número de mortes por COVID-19 apresentou uma evolução heterogênea ao longo do tempo, então, para cada período específico da pandemia podemos ter um taxa de crescimento e/ou um modelo de crescimento diferente. Devido a isto, também consideramos o ajuste de um modelo por partes. No ajuste deste modelo, consideramos como critério para identificação de um ponto de mudança o seguinte procedimento: (i) Seleção do melhor modelo (log-exponencial, log-logístico ou log-Gompertz) para os dados registrados em um determinado intervalo de tempo com d dias e construção de um intervalo de predição; (ii) se o intervalo não conter o valor registrado no dia $d + 1$, então, consideramos o dia $d + 1$ como sendo um ponto de mudança. Este procedimento é repetido de forma sequencial até chegarmos ao último dia do período considerado neste estudo (30/04/2021). Como resultado, identificamos seis períodos com os seguintes tipos de crescimento: Gompertz, exponencial, logístico, logístico, exponencial e exponencial.

As duas principais vantagens em considerar o ajuste de um modelo por partes são: 1. sua eficácia na descrição dos diferentes períodos da pandemia; e 2. sua capacidade de explicar cada período da pandemia por meio das estimativas para os parâmetros epidemiológicos de interesse, tais como, a taxa de crescimento do número de mortes diárias e a data de ocorrência do pico.

Uma vez que a modelagem proposta foi aplicada aos dados após o final da pandemia, informamos ao leitor que objetivo principal do artigo não é a apresentação de um modelo de predição, mas sim, a apresentação de um modelo de regressão capaz de explicar como se deu a evolução do logaritmo do número acumulado de mortes por COVID-19 no estado de São Paulo no período de 17/03/2020 a 30/04/2021, e apresentar as taxas de mortes e datas em que ocorreram os picos em cada período.

O restante do artigo está organizado da seguinte maneira. Na Seção 2, apresentamos uma revisão detalhada dos modelos de crescimento exponencial, logístico e de Gompertz. Na Seção 3, descrevemos um estudo de caso, em que, utilizamos os modelos de crescimento na modelagem do

número acumulado de mortes por COVID-19 registrados no estado de São Paulo no período de 17/03/2020 a 30/04/2021. No Seção 4, apresentamos as considerações finais.

2 MODELOS EXPONENCIAL, LOGÍSTICO E DE GOMPERTZ

Nesta seção, apresentamos uma revisão sobre os modelos de crescimento exponencial, logístico e de Gompertz.

2.1 Modelo exponencial

Quando um aluno do ensino médio é iniciado nos estudos sobre Biologia, este é apresentado ao conceito de mitose. De acordo com [1], a mitose é um tipo de divisão celular que ocorre em todas as células eucarióticas e garante a formação de duas células-filhas.

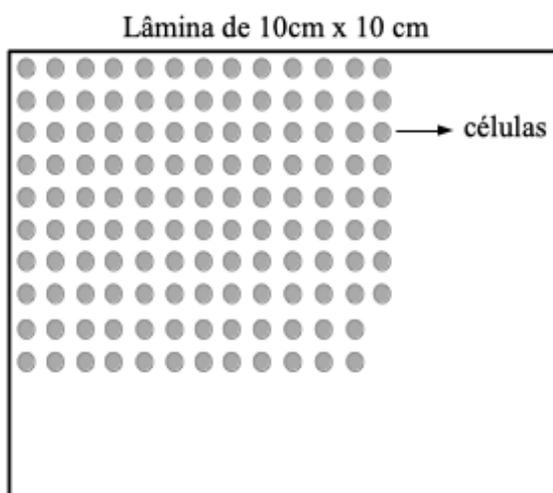


Figura 1: Divisão celular em 8 dias.

Agora, suponha que para entender esta dinâmica de crescimento, um matemático realiza este experimento por oito dias, utilizando uma lâmina de vidro quadrada de lado igual a 10cm, conforme ilustrado na Figura 1. Ou seja, no dia 0 (início do experimento), há uma única célula. No dia 1, haverá duas células, no dia 2 haverá 4 células e sucessivamente até o dia 7, em que, haverá 128 células. A pergunta que surge neste ponto é: se realizarmos este experimento por t dias, qual será o tamanho N_t desta população?

Sob estas condições, sendo N_0 o tamanho da população no dia 0, temos que, o tamanho da população em um dia t é o dobro do dia anterior, $t - 1$, *i.e.*,

$$\begin{aligned} N_1 &= 2N_0 &\Rightarrow N_1 &= 2N_0 \\ N_2 &= 2N_1 &\Rightarrow N_2 &= 2^2N_0 \\ N_3 &= 2N_2 &\Rightarrow N_3 &= 2^3N_0 \\ \vdots & \quad \vdots & \quad \vdots & \Rightarrow \quad \vdots \quad \vdots \quad \vdots \end{aligned}$$

Ou seja, temos que o tamanho da população em um instante de tempo t qualquer é dado por $N_t = 2^t N_0$, para $t > 0$. Neste caso, dizemos que a taxa de crescimento da população é 2, pois ela dobra a cada dia. Contudo, a expressão $N_t = 2^t N_0$ é válida apenas para a situação, em que, a taxa de crescimento é 2. Por exemplo, se a cada dia tivermos duas reproduções (*i.e.*, a formação de quatro células filhas), então a expressão é dada por $N_t = 4^t N_0$, pra $t > 0$.

Assim, de maneira geral, denotando a taxa de crescimento por r , então o tamanho da população em um instante de tempo t qualquer pode ser escrita como:

$$N_t = r^t N_0, \tag{2.1}$$

para $t > 0$. Este modelo é denominado na literatura de modelo exponencial.

A Figura 2, mostra o gráfico do modelo exponencial para $N_0 = 1$ e $r = \{2, 3, 4\}$. Note que, aumentando o valor de r , mais inclinada é a curva, *i.e.*, mais rápido é o crescimento populacional. Da Equação em (2.1), temos que $\frac{N_t}{N_0} = r^t$. Tomando o logaritmo de base e , em ambos os lados desta Equação, obtemos $\ln\left(\frac{N_t}{N_0}\right) = t \ln(r)$. Considerando $\alpha = \ln(r)$, temos que, $\ln\left(\frac{N_t}{N_0}\right) = t \alpha$ e conseqüentemente $\frac{N_t}{N_0} = e^{\alpha t}$. Assim, obtemos a seguinte forma de apresentar um modelo exponencial:

$$N_t = N_0 e^{\alpha t}, \tag{2.2}$$

onde $\alpha = \ln(r)$ é denominado de taxa intrínseca de crescimento por unidade de tempo, para $t > 0$.

Além disso, note que, derivando a Equação em (2.2) em relação a t , obtemos:

$$\frac{dN_t}{dt} = \alpha N_0 e^{\alpha t} = \alpha N_t, \tag{2.3}$$

onde $\frac{dN_t}{dt}$ é denominada de taxa de crescimento instantânea da população no ponto t , para $t > 0$. Ou seja, uma maneira alternativa de escrever o modelo exponencial descrito na Equação em (2.2) é através da equação diferencial em (2.3).

Do modelo em (2.3), temos que, a taxa de crescimento instantânea é sempre proporcional ao valor atual da função. Esta é uma característica chave do crescimento exponencial. Além disso, como $\alpha = \ln(r) > 0$, então, os valores de N_t crescem indefinidamente, conforme ilustrado no gráfico da Figura 2. Em outras palavras, este modelo não considera a existência de mecanismos que possam limitar o crescimento da população, tal como, recursos limitados de espaço ou alimentos, por exemplo.

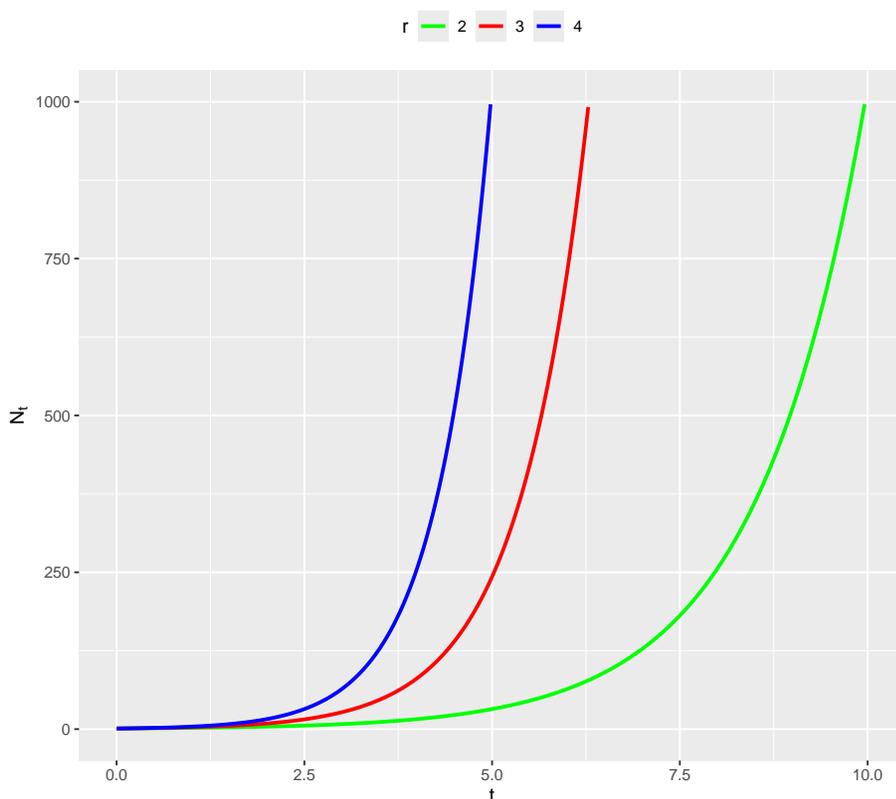


Figura 2: Modelo exponencial.

Do ponto de vista prático, este crescimento ilimitado do modelo exponencial, pode ser uma desvantagem, pois em diversos problemas temos algum tipo de limitador para o crescimento populacional. Assim, é de interesse termos um modelo de crescimento que leve em consideração a presença de mecanismos limitadores do crescimento populacional. Dois modelos com esta característica são: (i) o modelo logístico, proposto por [14], e (ii) o modelo de Gompertz, proposto por [5].

2.2 Modelo logístico

Para introduzirmos o modelo logístico, considere, sem perda de generalidade, a ilustração apresentada na Figura 1. Nesta Figura, o espaço de reprodução das células é uma lâmina quadrada de lado igual a 10 cm . Logo, o número de células que podemos ter sobre esta lâmina fica limitada a área de 100 cm^2 , fazendo com que haja um limitante no crescimento populacional das células.

Como no início do experimento há um número maior de recursos (neste exemplo, uma área maior para reprodução) e pouca população de células sobre a lâmina, é natural considerar que haja um crescimento populacional acelerado. No entanto, à medida que a população de células

se aproxima do limite da área total da lâmina, também é natural considerar que a capacidade de reprodução tenda a diminuir. Ou seja, temos uma taxa de crescimento que possui uma dependência com o tamanho populacional das células em um determinado instante de tempo t , para $t > 0$.

Sob este cenário, [2] informa que [14] propõe utilizar a taxa $(1 - \frac{N_t}{K})$, onde K é a quantidade máxima de células que se pode ter na placa. Dessa forma, no início do processo, em que, N_t é pequeno em relação a K , temos $(1 - \frac{N_t}{K}) \approx 1$, indicando uma taxa de crescimento alta. Por outro lado, à medida que N_t se aproxima do limitante K , temos $(1 - \frac{N_t}{K}) \approx 0$, indicando uma taxa de crescimento menor.

Então, [14] propõe a seguinte modificação do modelo exponencial:

$$\frac{dN_t}{dt} = \alpha N_t \left(1 - \frac{N_t}{K} \right), \tag{2.4}$$

para $t > 0$. Considerando a notação $\frac{dN_t}{dt} = N'_t$, podemos escrever a Equação em (2.4) como:

$$N'_t = \alpha N_t - \frac{\alpha N_t^2}{K}, \tag{2.5}$$

para $t > 0$. Ou seja, temos uma equação diferencial ordinária (EDO) de Bernoulli [7].

Para obter a solução da EDO em (2.5), considere a seguinte mudança de variável $z = N_t^{-1}$. Assim, temos que, $N'_t = -z'z^{-2}$, onde $z' = \frac{dz}{dt}$. Logo, podemos reescrever a Equação (2.5) da seguinte maneira:

$$-z'z^{-2} = \alpha_1 z^{-1} - \alpha_2 z^{-2}, \tag{2.6}$$

onde $\alpha_1 = \alpha$ e $\alpha_2 = \frac{\alpha}{K}$.

Multiplicando ambos os lados da Equação (2.6) por z^2 , obtemos a EDO linear:

$$z' + \alpha_1 z = \alpha_2. \tag{2.7}$$

Utilizando o método do fator integrante para resolver esta EDO, temos que:

$$\mu = \exp \left\{ \int \alpha_1 dt \right\} = e^{\alpha_1 t} \quad \text{e} \quad e^{\alpha_1 t} z' + \alpha_1 e^{\alpha_1 t} z = \alpha_2 e^{\alpha_1 t}.$$

Como $(e^{\alpha_1 t} z)' = e^{\alpha_1 t} z' + \alpha_1 e^{\alpha_1 t} z$, temos que $(e^{\alpha_1 t} z)' = \alpha_2 e^{\alpha_1 t}$. Logo,

$$e^{\alpha_1 t} z = \int \alpha_2 e^{\alpha_1 t} dt = \alpha_2 \int e^{\alpha_1 t} dt = \frac{\alpha_2}{\alpha_1} e^{-\alpha_1 t} + c,$$

onde c é uma constante. Portanto, a solução geral da Equação em (2.7) é dada por:

$$z = \frac{\alpha_2}{\alpha_1} + ce^{-\alpha_1 t} = \frac{1}{K} + ce^{-\alpha t},$$

pois $\alpha_1 = \alpha$ e $\alpha_2 = \frac{\alpha}{K}$.

Assim, a solução geral para o tamanho populacional N_t é dada por:

$$N_t = \frac{1}{\frac{1}{K} + ce^{-\alpha t}}, \tag{2.8}$$

para $t > 0$. Geometricamente, esta função define uma família de curvas em forma de S, uma para cada valor da constante c , conforme ilustrado na Figura 3 para $K = 1.000$, $c = \{1, 10, 100\}$ e $\alpha = 0,10$.

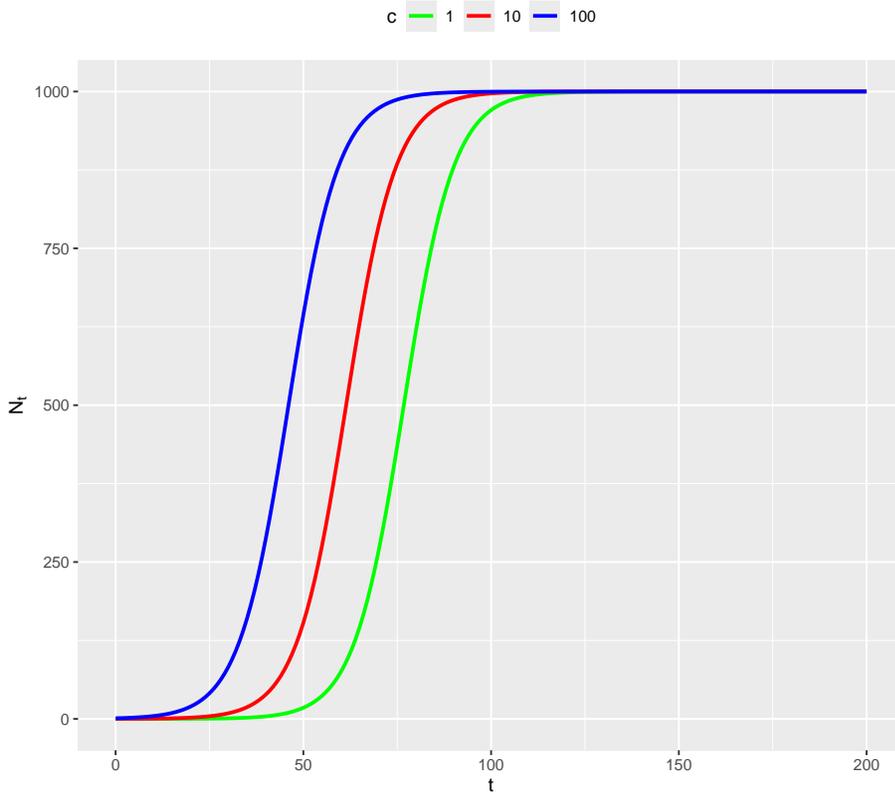


Figura 3: Modelo logístico.

Como o interesse é na curva que passa pelo ponto de coordenadas $(t = 0, N_0)$, vamos determinar o valor de c para este caso. Assim, temos que:

$$N_0 = \frac{1}{\frac{1}{K} + ce^{-\alpha_1 \cdot 0}} = \frac{1}{\frac{1}{K} + c}.$$

Isolando c , obtemos:

$$c = \frac{1}{N_0} - \frac{1}{K}.$$

Portanto, a solução da Equação diferencial dada em (2.4), com condição inicial $(t = 0, N_0)$, é dada por:

$$N_t = \frac{1}{\frac{1}{K} + \left(\frac{1}{N_0} - \frac{1}{K}\right)e^{-\alpha t}} = \frac{K}{1 + \beta e^{-\alpha t}}, \tag{2.9}$$

para $t > 0$, onde $\beta = c \cdot K = \left(\frac{K}{N_0} - 1\right)$. Esta expressão define o que na literatura é denominado de modelo logístico de parâmetros (K, β, α) .

2.2.1 Ponto de inflexão

Definida a função de crescimento N_t , há o interesse em determinar o ponto em que a taxa de crescimento atinge o valor máximo, denominado de ponto de inflexão. De acordo com [3]: “O ponto de inflexão de uma função contínua $f(t)$ é um ponto $t = p$, em um intervalo aberto contendo o ponto $t = p$, onde a segunda derivada $f''(t) > 0$ de um lado de $t = p$ e $f''(t) < 0$ do outro lado de $t = p$, e $f''(p) = 0$ ou não existe”.

Derivando a função em (2.9) em relação a t obtemos: $\frac{dN_t}{dt} = N'_t = \alpha N_t \left(1 - \frac{N_t}{K}\right)$. A segunda derivada é dada por:

$$\frac{d^2N_t}{dt^2} = N''_t = \alpha N'_t \left(1 - \frac{2N_t}{K}\right).$$

Logo, temos que $N''(t) = 0$ se e somente se: $\left(1 - \frac{2N_t}{K}\right) = 0 \Leftrightarrow N_t = \frac{K}{2}$. Isto implica que $t = p = \ln(\beta)/\alpha$ é candidato a ponto de inflexão.

Precisamos agora, verificar se $t = p$ satisfaz a condição de ponto de inflexão descrita por [3]. Note que, para $N''_t > 0$ devemos ter $\left(1 - \frac{2N_t}{K}\right) > 0$, pois $\alpha > 0$ e $N'_t > 0$. Logo, devemos ter $N_t < \frac{K}{2}$, implicando que $t < p$. Similarmente, para $N''_t < 0$ devemos ter $t > p$. Logo, as coordenadas do ponto de inflexão da curva de N_t é dada por $\left(\frac{\ln(\beta)}{\alpha}, \frac{K}{2}\right)$.

Como ilustração, a Figura 4(a) mostra o gráfico de N''_t para $K = 1.000$, $\beta = 400$ e $\alpha = 0,15$ destacando o ponto $(p, 0)$, sendo $p = \frac{\ln(400)}{0,15} = 39,9431$. Note que, $N''_t > 0$ para valores $t < p$ e $N''_t < 0$ para valores $t > p$. A Figura 4(b) mostra o gráfico do modelo logístico para $K = 1.000$, $\beta = 400$ e $\alpha = \{0,10; 0,15; 0,20\}$ destacando o ponto de inflexão (símbolo ●). Note que, aumentando o valor do parâmetro α , maior é a inclinação da curva.

2.3 Modelo de Gompertz

Similar ao modelo logístico, o modelo de Gompertz [5], também considera a existência de mecanismos limitantes. Análogo à descrição do modelo logístico, vamos obter o modelo de Gompertz a partir do modelo exponencial. Para isto, considere o modelo exponencial dado na Equação (2.3).

Suponha que a taxa de crescimento α dependa do tempo t , i.e., $\alpha = \alpha_t$, para $t > 0$. Assim, temos que:

$$\frac{dN_t}{dt} = \alpha_t N_t, \tag{2.10}$$

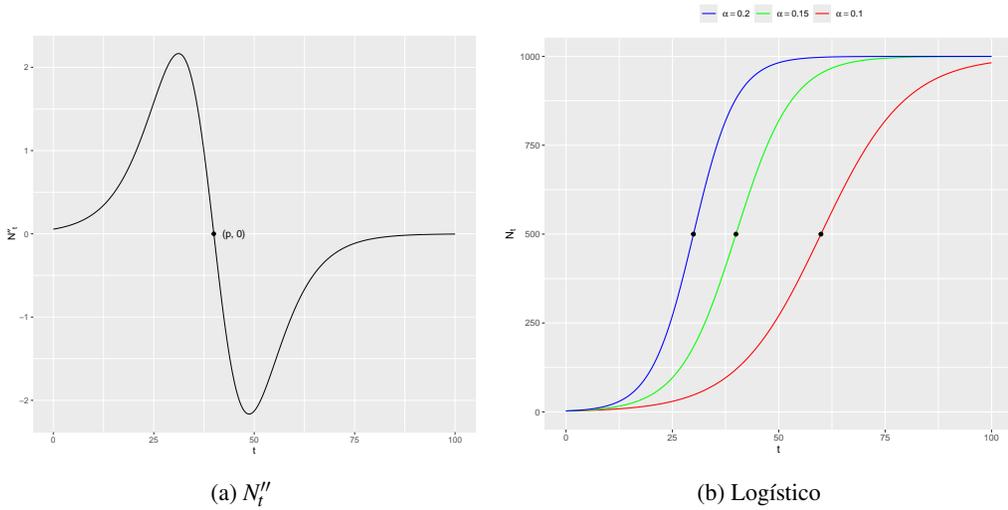


Figura 4: Gráfico de N_t'' e do modelo logístico destacando o ponto de inflexão.

para $t > 0$. Além disso, considere que taxa de crescimento α_t decresce de maneira exponencial em função do tempo, *i.e.*, $\frac{d\alpha_t}{dt} = -a\alpha_t \Rightarrow \alpha_t = \frac{-1}{a} \frac{d\alpha_t}{dt}$, onde a é um constante positiva, para $t > 0$.

Da expressão em (2.10), temos que, $\alpha_t = \frac{1}{N_t} \frac{dN_t}{dt}$. Além disso, temos que:

$$\frac{d(\ln(N_t))}{dt} = \frac{1}{N_t} \frac{dN_t}{dt} = \alpha_t = \frac{-1}{a} \frac{d\alpha_t}{dt}.$$

Logo,

$$\ln(N_t) = (-\alpha_t + b)/a, \tag{2.11}$$

onde b é uma constante. Desta expressão, temos que, para $\alpha_t \downarrow 0$, $N_t \uparrow e^{b/a}$. Ou seja, a capacidade máxima desta população é:

$$K = e^{b/a} \Rightarrow b = a \ln(K). \tag{2.12}$$

De (2.11) e (2.12), temos que:

$$\alpha_t = b - a \ln(N_t) \Rightarrow \alpha_t = a \ln(K) - a \ln(N_t) \Rightarrow \alpha_t = a \ln\left(\frac{K}{N_t}\right). \tag{2.13}$$

Substituindo (2.13) em (2.10), temos que:

$$\frac{dN_t}{dt} = aN_t \ln\left(\frac{K}{N_t}\right); \tag{2.14}$$

e podemos escrever a Equação em (2.14) como:

$$\frac{dN_t}{N_t \ln\left(\frac{K}{N_t}\right)} = a dt. \tag{2.15}$$

Ou seja, temos uma equação diferencial ordinária (EDO) separável.

Integrando ambos os lados de (2.15) obtemos:

$$\int \frac{1}{N_t \ln\left(\frac{K}{N_t}\right)} dN_t = \int a dt. \tag{2.16}$$

Para obter a solução de (2.16), considere a seguinte mudança de variável: $u = \ln\left(\frac{K}{N_t}\right)$. Assim, temos que, $du = \frac{-dN_t}{N_t}$. A integral do lado esquerdo da igualdade na expressão (2.16) é dada por:

$$\int \frac{1}{N_t \ln\left(\frac{K}{N_t}\right)} dN_t = - \int \frac{du}{u} = -\ln(u) = -\ln\left[\ln\left(\frac{K}{N_t}\right)\right] + c_1,$$

onde c_1 é uma constante. Para a integral do lado direito da igualdade na expressão (2.16), sabemos que:

$$\int a dt = at + c_2,$$

onde c_2 é uma constante. Com isso, a solução de (2.16) é dada por:

$$-\ln\left[\ln\left(\frac{K}{N_t}\right)\right] + c_1 = at + c_2 \Rightarrow -\ln\left[\ln\left(\frac{K}{N_t}\right)\right] + c_3 = at, \tag{2.17}$$

para $c_3 = c_1 - c_2$.

De forma análoga ao modelo logístico, nosso interesse é na curva que passa pelo ponto $(t = 0, N_0)$. Dessa forma, determinemos o valor de c_3 para o qual isto ocorre. Então, temos que:

$$-\ln\left[\ln\left(\frac{K}{N_0}\right)\right] + c_3 = a \cdot 0 \Rightarrow c_3 = \ln\left[\ln\left(\frac{K}{N_0}\right)\right]. \tag{2.18}$$

Substituindo (2.18) em (2.17), segue que:

$$-\ln\left[\ln\left(\frac{K}{N_t}\right)\right] + \ln\left[\ln\left(\frac{K}{N_0}\right)\right] = at \Rightarrow \ln\left[\frac{\ln\left(\frac{K}{N_0}\right)}{\ln\left(\frac{K}{N_t}\right)}\right] = at. \tag{2.19}$$

Ou seja, $\ln\left(\frac{K}{N_0}\right) = \ln\left(\frac{K}{N_t}\right) e^{at}$. Fazendo $\beta = \ln\left[\frac{K}{N_0}\right]$ e isolando N_t , obtemos como solução da EDO em (2.15):

$$N_t = Ke^{-\beta e^{-at}}, \tag{2.20}$$

para $t > 0$. Esta expressão define o que na literatura é denominado de modelo de Gompertz de parâmetros (K, β, a) .

2.3.1 Ponto de Inflexão

Definida a função de crescimento N_t de Gompertz, vamos agora determinar seu ponto de inflexão. Assim, derivando (2.20) em relação a t obtemos:

$$\frac{dN_t}{dt} = aKe^{-\beta e^{-at}} \beta e^{-\beta e^{-at}} = aN_t \beta e^{-\beta e^{-at}}.$$

Neste ponto, é importante notar que:

$$\ln\left(\frac{K}{N_t}\right) = \ln(K) - \ln(N_t) = \ln(K) - \ln(K) + \beta e^{-\beta e^{-at}} = \beta e^{-\beta e^{-at}}.$$

Logo, temos que: $N_t' = \frac{dN_t}{dt} = aN_t \ln\left(\frac{K}{N_t}\right)$. Assim, a segunda derivada de N_t em relação a t é dada por:

$$N_t'' = \frac{d^2N_t}{dt^2} = aN_t' \ln(K) - aN_t' \ln(N_t) - aN_t' = aN_t' \left[\ln\left(\frac{K}{N_t}\right) - 1 \right].$$

Logo, temos que $N_t'' = 0$ se e somente se:

$$\left[\ln\left(\frac{K}{N_t}\right) - 1 \right] = 0 \Rightarrow \ln\left(\frac{K}{N_t}\right) = 1 \Rightarrow N_t = \frac{K}{e}.$$

Isto implica que $t = p = \ln(\beta)/a$ é candidato a ponto de inflexão. Precisamos agora, verificar se $t = p$ satisfaz a condição de ponto de inflexão descrita por [3]. Note que, para $N_t'' > 0$ devemos ter $\ln\left(\frac{K}{N_t}\right) - 1 > 0$, pois $a > 0$ e $N_t' > 0$. Logo, devemos ter $N_t < \frac{K}{e}$, implicando que $t < p$. Similarmente, para $N_t'' < 0$ devemos ter $t > p$. Logo, as coordenadas do ponto de inflexão da curva de N_t é dada por $\left(\frac{\ln(\beta)}{a}, \frac{K}{e}\right)$. Ou seja, o ponto de inflexão ocorre para aproximadamente 36,79% da capacidade total K ; enquanto que para o modelo logístico o ponto de inflexão ocorre para metade da capacidade total K .

A Figura 5(a) mostra o gráfico de N_t'' para $K = 1.000$, $\beta = 400$ e $a = 0,15$ destacando o ponto $(p, 0)$, sendo $p = \frac{\ln(400)}{0,15} = 39,9431$. Note que, $N_t'' > 0$ para valores $t < p$ e $N_t'' < 0$ para valores $t > p$. A Figura 5(b) mostra o gráfico do modelo de Gompertz para $K = 1.000$, $\beta = 400$ e $a = \{0, 10; 0, 15; 0, 20\}$ destacando o ponto de inflexão (símbolo ●). Para este exemplo, a capacidade máxima é $K = 1.000$ e aumentando o valor do parâmetro a maior é a inclinação da curva.

2.4 Padronização da notação

Com o objetivo de simplificar a notação e torná-la padrão para desenvolvermos a aplicação dos três modelos de crescimento na modelagem do número acumulado de mortes por COVID-19, considere a seguinte parametrização para os modelos de crescimento:

- Exponencial: $N_t = h(t|\theta) = \alpha_1 e^{\alpha_2 t}$, para $\theta = (\alpha_1, \alpha_2)$ e $t > 0$, onde θ são os parâmetros do modelo.
- Logístico: $N_t = h(t|\theta) = \frac{\alpha_1}{1 + \alpha_2 \exp\{-\alpha_3 t\}}$, para $\theta = (\alpha_1, \alpha_2, \alpha_3)$ e $t > 0$, onde θ são os parâmetros do modelo.
- Gompertz: $N_t = h(t|\theta) = \alpha_1 \exp\{-\alpha_2 \exp\{-\alpha_3 t\}\}$, para $\theta = (\alpha_1, \alpha_2, \alpha_3)$ e $t > 0$; onde θ são os parâmetros do modelo.

Contudo, como o número acumulado de mortes por COVID-19, denotado por N_t , possuem valores “altos” (na escala de 1.000), isto pode causar problemas computacionais no processo de

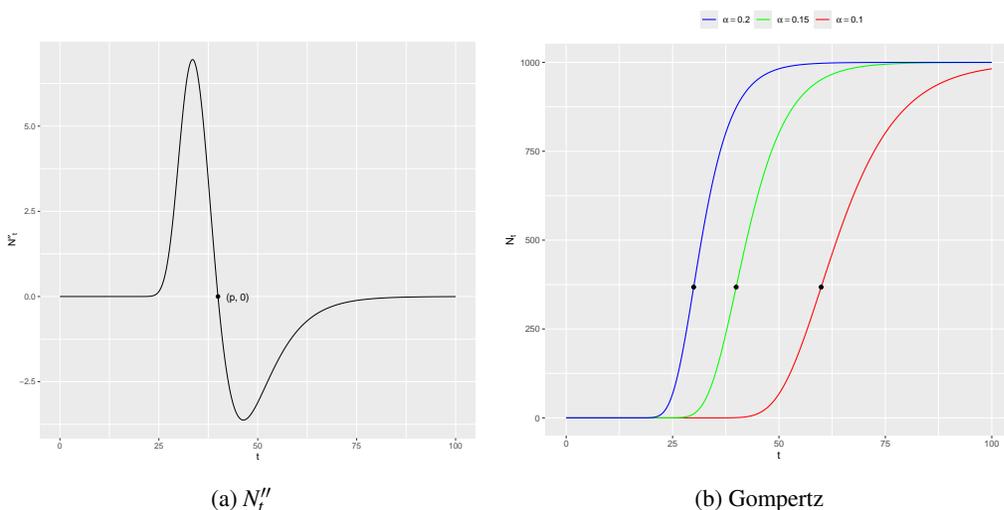


Figura 5: Gráfico de N_t'' e do modelo Gompertz destacando o ponto de inflexão.

estimação dos parâmetros dos modelos considerados. Assim, para evitar este problema, tomamos o logaritmo (base e) em ambos os lados de cada modelo e obtemos os modelos de crescimento:

- log-exponencial: $\ln(N_t) = \ln(h(t|\theta)) = \ln(\alpha_1) + \alpha_2 t$;
- log-logístico: $\ln(N_t) = \ln(h(t|\theta)) = \ln(\alpha_1) - \ln(1 + \alpha_2 \exp\{-\alpha_3 t\})$;
- log-Gompertz: $\ln(N_t) = \ln(h(t|\theta)) = \ln(\alpha_1) - \alpha_2 \exp\{-\alpha_3 t\}$;

para $t > 0$, em que θ são os parâmetros dos modelos.

A Figura 6 ilustra o gráfico dos três modelos na escala original e na escala logarítmica para $\alpha_1 = 1$ e $\alpha_2 = 0,15$ no modelo exponencial; e $\alpha_1 = 1.000$, $\alpha_2 = 700$ e $\alpha_3 = 0,1$ nos modelos logístico e de Gompertz.

3 APLICAÇÃO

Considere agora a aplicação dos modelos de crescimento exponencial, logístico e de Gompertz, na modelagem dos dados referente ao número acumulado de mortes devido a COVID-19, registrados no estado de São Paulo, Brasil, no período de 17/03/2020 (registro do primeiro óbito) a 27/04/2021 (último dia considerado neste estudo). Neste período de 410 dias foram registrados um total de 96.191 mortes.

Para iniciarmos o procedimento de modelagem, denote por M_t o número de óbitos por COVID-19 registrados no t -ésimo dia, para $t = 0, \dots, n$, onde $t = 0$ representa o dia em que foi registrado o primeiro óbito e $n = 410$ é o último dia considerado no estudo. Seja N_t a quantidade acumulada de mortes registradas até o t -ésimo dia, *i.e.*, $N_t = \sum_{t'=1}^t M_{t'}$, para $t = 0, \dots, n$. A Figura 7, mostra

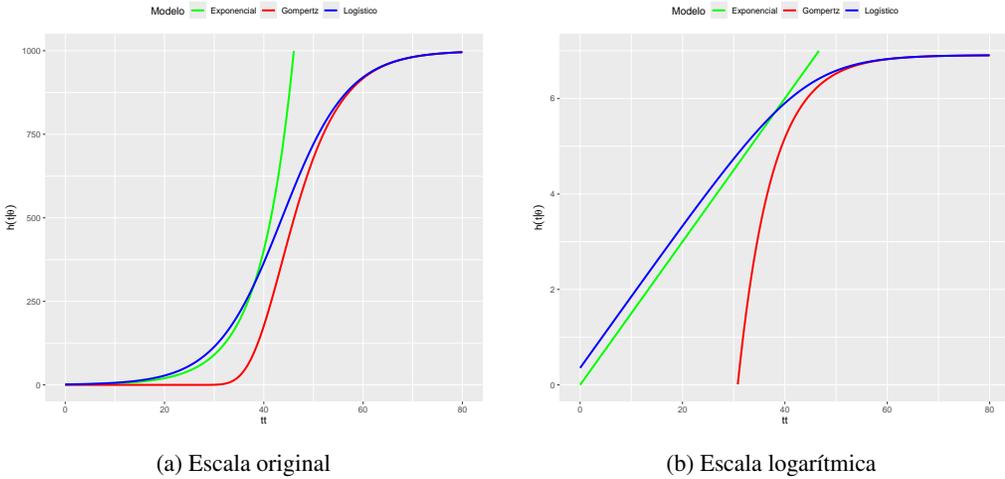


Figura 6: Gráficos dos modelos na escala original e na escala logarítmica.

o gráfico dos valores M_t e N_t , para $t = 0, \dots, n$. Note que, os valores N_t apresentam um comportamento mais estável do que os valores M_t ; tornando a modelagem de N_t mais simples do que a modelagem dos valores M_t .

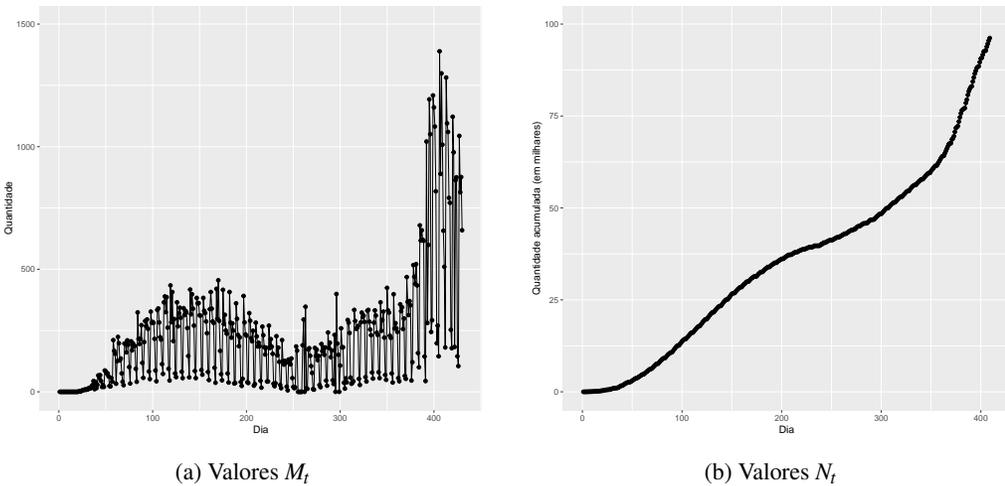


Figura 7: Período de 17/03/2020 a 30/04/2021.

No entanto, como muitos valores de N_t estão na escala de 1.000, isto pode causar problemas computacionais no processo de estimação dos parâmetros dos modelos considerados. Além disso, para ser possível utilizar a suposição de normalidade para os erros aleatórios, usual na mode-

lagem estatística, é preciso que os dados modelados sejam do tipo contínuo. Assim, considere $Y_t = \ln(N_t)$ e assuma que estes valores são gerados de acordo com o seguinte modelo aditivo:

$$Y_t = \ln(N_t) = f(t|\theta) + \varepsilon_t, \tag{3.1}$$

onde $f(t|\theta) = \ln(h(t|\theta))$ é a função de um dos modelos de crescimento da escala log e ε_t é um erro aleatório assumido como sendo gerado de uma distribuição normal com média $\mu = 0$ e variância σ^2 , i.e., $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, para $t > 0$.

3.1 Estimação dos parâmetros

Definido o modelo estatístico em (3.1), nosso principal interesse é obter estimativas para os parâmetros θ do modelo considerado com base nos dados observados $\mathbf{y} = (y_1, \dots, y_n)$. Ao obter a estimativa $\hat{\theta}$ de θ , temos o modelo ajustado: $\hat{Y}_t = f(t|\hat{\theta})$, onde \hat{Y}_t é o valor estimado pelo modelo para Y_t , $t > 0$.

A diferença entre o valor observado Y_t e o valor estimado \hat{Y}_t é denominado de resíduo. Vamos denotar o resíduo para a t -ésima observação por e_t , i.e., $e_t = Y_t - \hat{Y}_t$, para $t > 0$. Ou seja, o resíduo e_t é o valor observado do erro aleatório ε_t , para $t > 0$.

Assim, para o processo de estimação dos parâmetros θ , é natural requerer que a curva ajustada aos dados seja tal que o valor e_t seja o menor possível, para todo $t > 0$. Ou seja, se observamos (Y_1, \dots, Y_n) nos instantes de tempo $(t_1, \dots, t_n) = (1, \dots, n)$, então desejamos obter uma curva, tal que, o comprimento do vetor $\mathbf{e} = (e_1, \dots, e_n)$ seja mínimo. Utilizando a norma Euclideana, isto equivale a minimizar a medida denominada de soma dos quadrados dos resíduos (*SQRes*), definida como

$$SQRes = \sum_{t=1}^n e_t^2 = \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 = \sum_{t=1}^n (Y_t - f(t|\hat{\theta}))^2.$$

Como este método minimiza o quadrado das distâncias, ele é denominado na literatura de método dos mínimos quadrados [15].

3.1.1 Modelo log-exponencial

Para o modelo log-exponencial, $SQRes = \sum_{i=1}^n (Y_t - \alpha_1 - \alpha_2 t)^2$. Derivando *SQRes* em relação a α_1

e α_2 e igualando as expressões a zero, obtemos os estimadores: $\hat{\alpha}_1 = \bar{Y} - \hat{\alpha}_2 \bar{t}$ e $\hat{\alpha}_2 = \frac{\sum_{t=1}^n tY_t - n\bar{t}\bar{Y}}{\sum_{t=1}^n t^2 - n\bar{t}^2}$,

onde \bar{Y} e \bar{t} são as médias dos valores Y 's e t 's.

3.1.2 Modelo log-logístico

Para o modelo log-logístico, $SQRes = \sum_{i=1}^n [Y_t - \ln(\alpha_1) + \ln(1 + \alpha_2 e^{-\alpha_3 t})]^2$. Derivando $SQRes$ em relação a α_1 , α_2 e α_3 e igualando as expressões a zero, obtemos:

$$\begin{aligned} \frac{dSQRes}{d\alpha_1} &= \sum_{t=1}^n -\frac{2}{\alpha_1} [Y_t - \ln(\alpha_1) + \ln(1 + \alpha_2 e^{-\alpha_3 t})] = 0; \\ \frac{dSQRes}{d\alpha_2} &= \sum_{t=1}^n \frac{2e^{-\alpha_3 t}}{1 + \alpha_2 e^{-\alpha_3 t}} [Y_t - \ln(\alpha_1) + \ln(1 + \alpha_2 e^{-\alpha_3 t})] = 0; \\ \frac{dSQRes}{d\alpha_3} &= \sum_{t=1}^n -\frac{2t\alpha_2 e^{-\alpha_3 t}}{1 + \alpha_2 e^{-\alpha_3 t}} [Y_t - \ln(\alpha_1) + \ln(1 + \alpha_2 e^{-\alpha_3 t})] = 0; \end{aligned}$$

Contudo, não é possível isolar os parâmetros de interesse nas expressões acima. Ou seja, não é possível obter os estimadores para os parâmetros de interesse de maneira explícita. Devido a isto, precisamos obter as estimativas para os parâmetros de maneira numérica. Neste texto, utilizamos o software R e o comando `nls()` do pacote `nlstools` [4, 10]; pois este comando determina as estimativas de mínimos quadrados para os parâmetros de um modelo não linear.

3.1.3 Modelo log-Gompertz

Para o modelo log-Gompertz, $SQRes = \sum_{i=1}^n (Y_t - \ln(\alpha_1) + \alpha_2 e^{-\alpha_3 t})^2$. Derivando $SQRes$ em relação a α_1 , α_2 e α_3 e igualando as expressões a zero, obtemos

$$\begin{aligned} \frac{dSQRes}{d\alpha_1} &= \sum_{t=1}^n -\frac{2}{\alpha_1} [Y_t - \ln(\alpha_1) + \alpha_2 e^{-\alpha_3 t}] = 0; \\ \frac{dSQRes}{d\alpha_2} &= 2e^{-\alpha_3 t} \sum_{t=1}^n [Y_t - \ln(\alpha_1) + \alpha_2 e^{-\alpha_3 t}] = 0; \\ \frac{dSQRes}{d\alpha_3} &= -2t\alpha_2 e^{-\alpha_3 t} \sum_{t=1}^n [Y_t - \ln(\alpha_1) + \alpha_2 e^{-\alpha_3 t}] = 0. \end{aligned}$$

Para este modelo, também não é possível obter os estimadores para os parâmetros de maneira explícita. Devido a isto, vamos obter as estimativas para os parâmetros de maneira numérica utilizando o comando `nls()` do software R.

3.2 Resultados

Nesta seção, apresentamos os resultados da aplicação dos três modelos de crescimento aos dados do número acumulado de mortes por COVID-19 registrados no estado de São Paulo no período de 410 dias (de 17/03/2020 à 27/04/2021).

A Tabela 1, mostra as estimativas pontuais e intervalares (95%) para os parâmetros dos três modelos. A Tabela 2 mostra os valores do erro quadrático médio (EQM) e os valores dos critérios de

seleção de modelos AIC e BIC. Estes valores foram calculados de acordo com as seguintes expressões: $EQM = \frac{1}{n} \sum_{i=0}^n (\hat{Y}_i - Y_i)^2$, $AIC = -2l(\hat{\theta}|\mathbf{y}) + 2q$ e $BIC = -2l(\hat{\theta}|\mathbf{y}) + q \ln(n)$ em que, $l(\hat{\theta}|\mathbf{y})$ é a função log-verossimilhança para $\theta = \hat{\theta}$ e q é a quantidade de parâmetros no modelo. Os valores AIC e BIC foram calculados utilizando os comandos AIC() e BIC() do software R. O modelo com menores valores EQM, AIC e BIC é então considerado o melhor modelo dentre os testados. Os três critérios indicam o modelo log-Gompertz como sendo o melhor modelo.

Tabela 1: Estimativas para os parâmetros dos modelos.

Parâmetros	Modelo					
	log-exponencial		log-logístico		log-Gompertz	
	Estimativa	I.C. (95%)	Estimativa	I.C. (95%)	Estimativa	I.C. (95%)
$\ln(\alpha_1)$	7,4121	(7,2050; 7,6192)	10,6751	(10,6145; 10,7361)	10,8420	(10,8018; 10,8826)
α_2	0,0118	(0,0109; 0,0127)	696,6588	(557,9218; 873,9093)	8,0320	(7,8699; 8,1964)
α_3	-	-	0,0614	(0,0614; 0,0719)	0,0193	(0,0186; 0,0200)

Tabela 2: Valores EQM, AIC e BIC.

Critério	Modelo		
	log-exponencial	log-logístico	log-Gompertz
EQM	1,1252	0,2309	0,0762
AIC	1214,94	569,13	115,51
BIC	1226,98	585,19	131,56

A Figura 8 mostra os valores registrados (símbolo ●) e as curvas dos modelos ajustados na escala logarítmica e original. Note que, os três modelos considerados não modelam bem os dados registrados. Neste ponto, conjecturamos que isto ocorreu devido o número acumulado de mortes por COVID-19 ter uma evolução heterogênea ao longo do tempo, *i.e.*, períodos de tempo com forma e/ou taxa crescimento diferentes. Logo, a adoção de uma única função $f(t|\theta)$ para explicar todo o período considerado pode não ser adequada. Assim, para superar esta questão, propomos o ajuste de um modelo por partes.

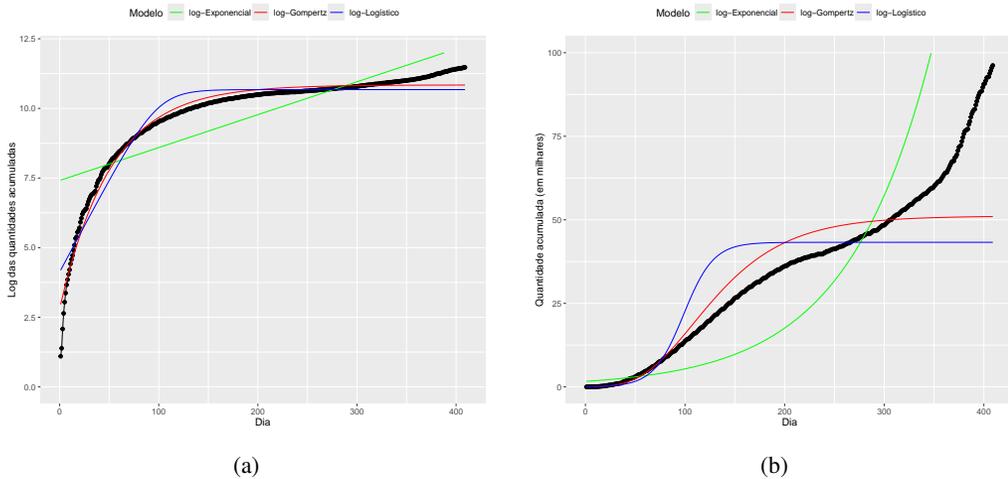


Figura 8: Período de 17/03/2020 a 30/04/2021.

3.3 Modelo por partes

Considere o modelo em (3.1) com $f(t|\theta)$ dado pela seguinte função por partes:

$$f(t|\theta) = \begin{cases} f_1(t|\theta) & , \text{ for } 0 < t \leq d_1; \\ \vdots & , \quad \vdots \quad \quad \quad ; \\ f_k(t|\theta) & , \text{ for } d_{k-1} < t \leq n; \end{cases}$$

onde d_1, \dots, d_{k-1} representam os pontos de mudança, ou seja, estamos assumindo que o período de $n = 410$ dias esta dividido em k períodos, sendo $f_j(t|\theta)$ a função que melhor explica os dados do período j , para $j = 1, \dots, k$. Consideramos, para $f_j(t|\theta)$, as funções dos modelos log-exponencial (LE), log-logístico (LL) e log-Gompertz (LG), sendo o melhor modelo para o período, escolhido de acordo com os critérios *EQM*, *AIC* e *BIC*.

Para determinar os pontos de mudança, propomos o seguinte procedimento:

- (1) Inicie o conjunto de dados $D_1 = \{y_1, \dots, y_{20}\}$ com $n_1 = 20$. A escolha do valor $n_1 = 20$ é devido ao fato de que este é o valor mínimo que encontramos em um estudo “piloto” para que haja convergência do algoritmo utilizado para obter as estimativas para os parâmetros dos modelos LL e LG.
- (2) Obtenhas as estimativas para os parâmetros dos modelos LE, LL e LG.
- (3) Escolha o melhor modelo com base nos critérios *EQM*, *AIC* e *BIC*.
- (4) Utilizando o modelo selecionado construa um intervalo de $100(1 - \alpha\%)$ de predição para o valor y_{n_1+1} , para $0 < \alpha < 1$. Para obter este intervalo, utilizamos o software R e o comando `predFit()` com a opção `interval='prediction'`. Fixamos $\alpha = 0,10$, para obter um intervalo de predição de 90% confiança.

- (i) se o intervalo conter o valor y_{n_1+1} , faça $D_1 \cup \{y_{n_1+1}\}$, $n_1 = n_1 + 1$ e atualize as estimativas para os parâmetros do modelo selecionado.
 - (ii) se o intervalo não conter o valor y_{n_1+1} ; então, consideramos y_{n_1+1} como sendo um ponto de mudança. Neste caso, vá para o itens (5) e (6).
- (5) Inicie o conjunto de dados $D_2 = \{y_{n_1+1}, \dots, y_{n_1+19}\}$ com $n_2 = 20$. Repita os itens (1) a (4) para D_2 até encontrar um novo ponto de mudança.
- (6) Sempre que encontrar um ponto de mudança, inicie um novo conjunto de dados e repita os itens (1) a (5).

Aplicamos o procedimento descrito acima aos dados referente ao número acumulado de óbitos registrados no estado de São Paulo. No total foram identificados $k = 6$ fases conforme descrito na Tabela 3. Esta Tabela mostra o modelo, as estimativas para os parâmetros do modelo selecionado para cada fase, o período em que cada modelo ficou vigente e a quantidade de dias de cada fase.

Tabela 3: Fases identificadas e estimativas para os parâmetros.

Fase	Modelo	Estimativa para os parâmetros			Período	Quantidade de dias
		$\ln(\alpha_1)$	α_2	α_3		
F_1	log-Gompertz	7,8981	6,9219	0,0611	17/03/20 a 27/04/20	42
F_2	log-logístico	9,7537	7,0108	0,0522	28/04/20 a 18/06/20	52
F_3	log-logístico	10,6629	2,4613	0,0246	19/06/20 a 17/11/20	152
F_4	log-exponencial	10,6157	0,0030	-	18/11/20 a 17/12/20	30
F_5	log-exponencial	10,6964	0,0042	-	18/12/20 a 10/03/21	83
F_6	log-logístico	12,6049	3,8246	0,0123	11/03/21 a 30/04/21	51

A Figura 9 mostra o gráfico dos valores registrados (símbolos ●) e a curva do modelo ajustado (linha em vermelho) na escala log e na escala original, em que, F_j entre as linhas tracejadas indica o período da fase j , para $j = 1, \dots, 6$. De acordo com o modelo ajustado, a fase 1 apresentou um crescimento do tipo LG, com pico estimado para o dia 34 desta fase (19/04/2020) e estimativa para a quantidade máxima de mortes de 2.693. Contudo, nove dias após a ocorrência do “pico” da primeira fase identificamos uma mudança de comportamento dos valores registrados.

No dia 42 (28/04/20) teve início a segunda fase, sendo esta “nova” fase explicada pelo modelo LL; com pico estimado para ocorrer no dia 38 desta fase (04/06/20) e estimativa para a máxima quantidade de mortes de 17.218. Este comportamento do tipo LL permaneceu por 52 dias (até o dia 93, 18/06/20). No dia 94 (19/06/20) identificamos uma nova mudança de comportamento, o qual denominamos de fase 3. A fase 3 também apresentou um comportamento explicado pelo modelo LL, com “novo pico” estimado para ocorrer no dia 37 desta fase (25/07/20) e estimativa para a máxima quantidade de mortes de 42.741. Esta foi a maior fase identificada pela modelagem proposta, com duração de 152 dias (até o dia 245, 17/11/20).

No dia 246 (18/11/20) teve início a fase 4. Nesta fase, temos um crescimento explicado pelo modelo LE, que permaneceu até dia 275 (17/12/20). No dia 276 (18/12/20), ocorreu a mudança

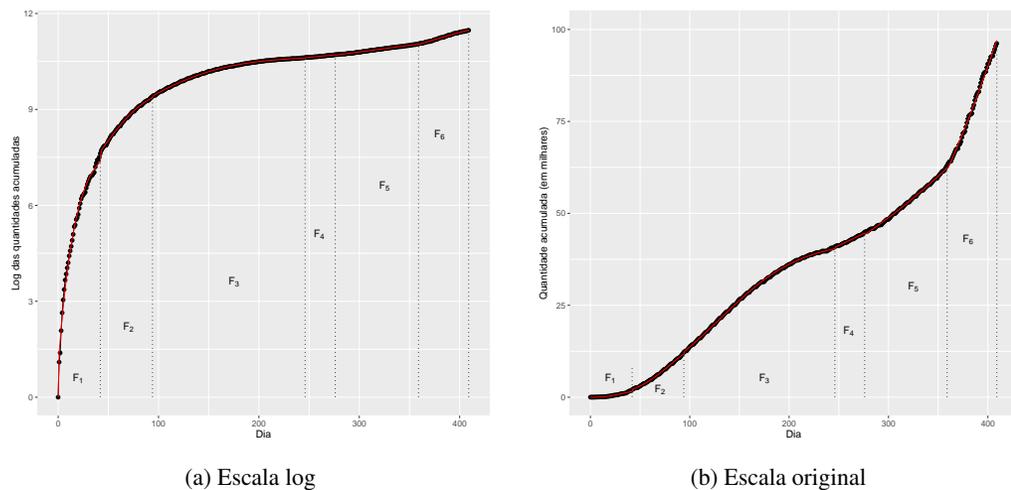


Figura 9: Período de 17/03/2020 a 30/04/2021.

para fase 5; também explicada por um modelo LE. Porém, a fase 5 apresenta um comportamento mais preocupante em relação a fase 4, pois a taxa de crescimento é maior do que do modelo LL da fase anterior (fase 4). No dia 359 ocorreu a mudança para a fase 6; que é explicada por um modelo LL com pico estimado para o dia 109 desta fase e estimativa para a quantidade máxima de mortes de 298.024.

Com relação ao ajuste do modelo, o EQM do modelo por partes foi de 0,0018. Ou seja, o modelo por partes é preferível em relação ao ajuste somente do modelo log-Gompertz (ver Tabelas 1 e 2), pois apresenta menor valor de EQM.

4 CONSIDERAÇÕES FINAIS

Neste artigo, apresentamos uma revisão dos modelos de crescimento exponencial, logístico e Gompertz. Descrevemos a utilização deste três modelos em um contexto estatísticos, *i.e.*, quando os dados observados são gerados de acordo com um modelo de regressão, em que, o valor da função do modelo de crescimento dá o valor esperado da variável aleatória de interesse. Como método de estimação dos parâmetros dos modelos, adotamos o método dos mínimos quadrados não-linear. Apresentamos uma aplicação na modelagem do número acumulado de mortes por COVID-19 registrados no estado de São Paulo no período de 17/03/2020 a 30/04/2021.

Contudo, como a evolução do número acumulado de mortes apresentou um comportamento heterogêneo ao longo do tempo, o ajuste de um único modelo de crescimento não apresentou resultados satisfatórios (ver Figura 8). Devido a isto, propomos o ajuste de um modelo por partes com pontos de mudanças determinados pela construção de intervalos de predição (90%). Utilizando esta proposta, identificamos $k = 6$ fases de crescimento distintas (ver Figura 9). Comparado ao ajuste de somente um modelo de crescimento, o ajuste de um modelo por partes é preferível, pois

apresenta menor valor de EQM . Além disso, o ajuste de um modelo por partes tem a vantagem de mostrar as diferentes fases de evolução do número de mortes devido a doença ao longo do tempo. Os códigos computacionais utilizados para ajuste dos modelos estão na linguagem R e podem ser obtidos via e-mail aos autores.

REFERÊNCIAS

- [1] A. Bruce, A. Johnson, J. Lewis, M. Raff, K. Roberts & P. Walter. “Biologia Molecular da Célula 5 ed”. Artmed, Porto Alegre: Artmed (2010).
- [2] J.S. Cramer. The early origins of the logit model. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, **35** (2004), 613–626.
- [3] C.H. Edwards & D.E. Penny. “Calculus with analytic geometry”. Printice Hall International, New Jersey (1993).
- [4] B. Florent, C. Ritz, M. Charles, J.P. Brustsche, P. Fladros & M.L. Delignette-Muller. A toolbox for nonlinear regression in R: The package nlstools. *Journal of Statistical Software*, **66** (2015), 1–21.
- [5] B. Gompertz. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London*, **115** (1825), 513–585.
- [6] P. Gosh, R. Gosh & B. Chakraborty. COVID-19 in India: Statewise Analysis and Prediction. *JMIR Public Health Surveill*, **26** (2020), e20341.
- [7] E. Hairer, S.P. Norsett & G. Wanner. “Solving ordinary differential equations I: Nonstiff problems”. Springer-Verlag, New York (1993).
- [8] C. Mahanty, R. Kumar, B.K. Mishra, D.J. Hemanth, D. Gupta & a. Khanna. Prediction of COVID-19 active cases using exponential and non-linear growth models. *Expert systems*, **39** (2021), e12648.
- [9] S.S. Musa, S. Zhao & M.H. Wang. Estimation of exponential growth rate and basic reproduction number of the coronavirus disease 2019 (COVID-19) in Africa. *Infectious Diseases of Poverty*, **9** (2020), 1225–1228.
- [10] J. Pinheiro, D. Bates, S. Debroy, D. Sakar & R Core Team. Linear and Nonlinear Mixed, R package version 3.1-147. URL <https://CRAN.R-project.org/package=nlme>, **6** (2020), 1–147.
- [11] A. Remuzzi & G. Remuzzi. COVID-19 and Italy: what next? *The lancet*, **397** (1825), 1225–1228.
- [12] E.F. Saraiva & C.A.B. Pereira. Piecewise Modeling the Accumulated Daily Growth of COVID-19 Deaths: The Case of the State of São Paulo, Brazil. *Entropy*, **23** (2021), 1013.
- [13] D. Sigh & S.V. Yi. On the origin and evolution of SARS-COV-2. *Experimental & Molecular Medicine*, **53** (2021), 537–546.
- [14] P.F. Verhulst. Deuxième mémoire sur la loi d’accroissement de la population. *Mémoires de l’Académie Royale des Sciences, des Lettres et des Beaux-Arts de Belgique*, **20** (1847), 1–32.
- [15] S. Vieira & R. Hoffman. Comparison of the Logistic and the Gompertz growth functions considering additive and multiplicative error terms. *Applied Statistics*, **6** (1977), 143–148.

ABSTRACT. The pandemic scenario caused by the SARS-CoV-2 coronavirus has increased the interest and divulgation of mathematical models capable of projecting the evolution of the number of cases (and/or deaths) due to COVID-19, in countries, states and/or cities. In many articles, the cumulative number of cases is modeled by a non-linear growth model, such as the exponential, logistic or Gompertz model. Motivated by this fact, in this article, we present a detailed review of these three growth models. We begin by obtaining the mathematical expression of the exponential model using a simple example of cell division. Based on the exponential model, we present in detail the mathematical development to obtain the expressions of the logistic and Gompertz models and how to obtain the coordinates of the inflection point of these two models. We also illustrate the use of these three growth models in the modeling of the cumulative number of deaths from COVID-19 recorded in the state of São Paulo in the period from 03/17/2020 to 04/30/2021. In this modeling, we present a criterion for adjusting a piecewise model since the recorded values showed heterogeneous behaviour over time.

Keywords: exponential model, logistic model, Gompertz model, inflexion point, statistical modeling, estimation.

Acknowledgements

Os autores agradecem a UFMS por todo o apoio. O autor W. A. O. SOLER agradece a FAPESP (grant 2024/07482-5) pelo apoio financeiro. O Autor E. F. Saraiva agradece ao CNPq (Processo 402154/2023-1) e a Fundect (TO 120/2024, SIAFIC 818, processo 83/026.835/2024) pelo apoio financeiro. Os autores agradecem aos revisores pelos comentários e sugestões dadas durante o processo de revisão.

Data availability

Datasets related to this article are available upon request to the corresponding author.

Associate Editor: Rosana Sueli Motta Jafelice

How to cite

Guimarães A.S., Rossi M.I.C.M., Soler W.A.O. & Saraiva E.F. Modelos de crescimento exponencial, logístico e Gompertz: Uma revisão teórica e aplicação a dados da COVID-19. *Trends in Computational and Applied Mathematics*, **26**(2025), e01837. doi: 10.5540/tcam.2025.026.e01837.

