



UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E COMPUTAÇÃO CIENTÍFICA
DEPARTAMENTO DE MATEMÁTICA APLICADA



SABRINA CAMARGO ZANI

Análise automática dos casos de COVID-19: Modelo SIR

Campinas
11/01/2021

SABRINA CAMARGO ZANI

Análise automática dos casos de COVID-19: Modelo SIR

Monografia apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos para obtenção de créditos na disciplina Projeto Supervisionado, sob a orientação do Prof. Alberto Saa.

Resumo

Esse trabalho desenvolveu-se em meio a pandemia mundial de COVID-19 e teve como objetivo assistir ao professor orientador em seu projeto de criação de um *website* que fizesse uma análise automática dos dados da pandemia no Brasil. Os dados utilizados foram os disponibilizados pelo próprio Ministério da Saúde na página oficial sobre COVID-19 do ministério e a análise estatística foi feita com base no modelo SIR. Nesse texto, será tratado especificamente o modelo SIR aplicado ao projeto.

Abstract

This work was carried out in the context of the global pandemic of COVID-19, with the aim to assist the professor in his project to develop a website that would make an automatic analysis of the Brazil's pandemic data. The data used were those published by the Ministry of Health of Brazil on its official webpage about COVID-19 and the statistical analysis was based on the SIR endemic model. In this text, we will talk specifically about the SIR model applied to the project.

Conteúdo

1	Introdução	1
2	O modelo SIR e a aplicação no projeto desenvolvido	2
2.1	Estrutura matemática do modelo	2
2.2	Parâmetro r_0	3
2.3	Parâmetro α	4
2.4	Estimativa dos parâmetros	5
3	Resultados	7
4	Conclusão	8

1 Introdução

Ao longo da história, populações de diferentes localizações foram acometidas por doenças infecciosas que vieram a tornar-se epidemias. A necessidade de estudo das doenças para o controle e erradicação de seus surtos fica clara então. Assim, a criação de modelos matemáticos epidemiológicos vem como uma ferramenta de grande ajuda nesse estudo. Esses modelos levam em conta as características de propagação, contaminação, imunização, entre outros das doenças a serem estudadas, colocando de forma exata os resultados experimentais encontrados durante os estudos da infecção. Dessa maneira, eles permitem fazer simulações e estimativas de como a epidemia se comportará, para que medidas possam ser tomadas para deter seu curso.

No ano de 2020, o mundo se deparou com uma pandemia de um vírus, o SARS-CoV-2, da família coronavírus que causa a doença infecciosa COVID-19. Esse vírus é a mutação de outro vírus da mesma família para que este possa infectar humanos. Ele foi primeiro identificado em Wuhan, China, em dezembro de 2019 e, mais tarde, espalhou-se por quase todo o mundo. Segundo o [4], a COVID-19 tem um espectro clínicos que variam desde infecções assintomáticas até quadros graves que necessitam de internação imediata e podem levar à morte.

Nessas condições que se deu o projeto do site [11] que visa fazer uma análise estatística automática dos dados da pandemia de coronavírus no Brasil a partir do modelo matemático epidemiológico SIR. Os dados utilizados na análise são retirados do site [4], transformados em um arquivo *.csv* ("*Comma Separated Value*") e tratados em um programa na linguagem de programação *python* que gera então o arquivo *HTML* do site. A esse respeito, mais informações podem ser encontradas em [3], [5] e [10].

Esse texto tem o objetivo de introduzir o modelo SIR como o modelo utilizado como base da análise dos dados do projeto. Há primeiro uma introdução do modelo e de sua expressão em um sistema de equações diferenciais ordinárias, depois a interpretação de seus parâmetros e, por fim, a discussão da estimativa dos mesmos.

2 O modelo SIR e a aplicação no projeto desenvolvido

Proposto por McKendrick e Kernack em 1927, o modelo SIR é um modelo epidemiológico compartimental dado por equações diferenciais ordinárias. Nele, uma população fixa (N) é dividida em três categorias (compartimentos): suscetíveis (S), os indivíduos que têm possibilidade de serem contaminados pela doença em questão; infectados (I), os que estão contaminados; e removidos (R), aqueles que não têm possibilidade de se contaminar nem de contaminar outros, ou seja, que foram isolados, que já se curaram da doença e estão imunes ou que faleceram por conta dela. Mais tarde foram criadas versões mais complexas do modelo, com mais compartimentos. Neste trabalho foi utilizada a versão simples do modelo SIR para a análise dos casos de COVID-19.

2.1 Estrutura matemática do modelo

Matematicamente, o modelo é descrito pelo seguinte sistema de EDOs:

$$\frac{dS}{dt} = -\frac{\beta IS}{N}, \quad (1)$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I, \quad (2)$$

$$\frac{dR}{dt} = \gamma I \quad (3)$$

Usualmente o parâmetro β caracteriza a taxa de infecção da doença, ou seja, ele representa a conversão da classe S para a classe I . Quanto maior o seu valor, maior é a facilidade de transmissão da COVID-19. Ele depende de diversos fatores, dentre eles temos a evolução da patologia, contato físico e hábitos de higiene dos indivíduos da população e fatores climáticos do local da epidemia. O parâmetro γ representa a conversão da classe I para a classe R . O seu inverso γ^{-1} caracteriza o tempo no qual um indivíduo fica infectado antes de ser realocado para classe dos removidos. Ambos os parâmetros são positivos e têm dimensão $[T^{-1}]$. Uma simulação desse modelo pode ser encontrada em [12].

2.2 Parâmetro r_0

Para o estudo da epidemia, a equação 2 pode ser reescrita da seguinte forma:

$$\frac{dI}{dt} = (r_0 \frac{S}{N} - 1)\gamma I \quad (4)$$

Em que o parâmetro adimensional r_0 , que é dado por

$$r_0 = \frac{\beta}{\gamma} \quad (5)$$

é um dos mais importantes nesse tipo de modelo e é chamado de número básico reprodutivo. Ele pode ser entendido como o número médio de pessoas (em S) que uma pessoa doente (em I) contamina.

Note que r_0 dita o crescimento do número de casos. Assim, $r_0 = 1$ está associado a fases endêmicas da infecção, em que o número de infectados permanece constante no tempo. Já para $r_0 < 1$, temos $\frac{dI}{dt} < 0$, pois $\gamma I > 0$, ou seja, o número de infectados no tempo está decaindo monotonicamente e a epidemia se extingue. De fato, como $0 \leq \frac{S}{N} \leq 1$, $r_0 < 1 \Rightarrow r_0 \frac{S}{N} < 1$.

Para $r_0 \frac{S}{N} > 1$ têm-se $\frac{dI}{dt} > 0$, a epidemia está se expandindo. No entanto, quanto maior o número de infectados, menor o número de suscetíveis, portanto a fração $\frac{S}{N}$ tende a diminuir até chegar ao seu limiar

$$\frac{S}{N} = \frac{1}{r_0} \quad (6)$$

Nesse limiar, a derivada temporal dos infectados se anula, ou seja, não há novas infecções. Essa situação é chamada de "imunidade de rebanho" e ocorre quando as condições da população local impedem novas contaminações. Nesse caso:

$$n_R = \frac{R}{N} = 1 - \frac{1}{r_0} = \text{constante} \quad (7)$$

Está claro que o parâmetro n_R cresce com o parâmetro r_0 , assim como a velocidade de expansão da epidemia. Da mesma forma, quanto menor o valor do número básico reprodutivo, mais rápido a epidemia se extingue.

2.3 Parâmetro α

Seja N , a população, constante, $N = R + S + I$ implica que é necessário o estudo de apenas dois dos três compartimentos para um entendimento completo do modelo. Foram escolhidos R e I como variáveis dinâmicas para tal estudo.

Os dados utilizados fornecem $C(t)$, o número de casos em função do tempo. Porém, na situação da pandemia de COVID-19, sabe-se que $C(t)$ é o número de casos de pessoas que vieram a ser hospitalizadas ou acudiram um hospital com sintomas moderados. Esse número foi incluído em R , como indivíduos que receberam recomendação de estrita de isolamento e/ou acabaram internados, dessa maneira não mais podendo infectar outros indivíduos.

Para essa doença, temos que considerar também o número de casos assintomáticos ou de sintomas leves que não são incluídos em $C(t)$, para nenhum t . Segundo [2], a relação dos casos em C para os casos reais está entre 1 : 10 e 1 : 20, que é a hipótese que foi utilizada nesse trabalho. Assim, definimos o parâmetro α como

$$R = \alpha C \quad (8)$$

admitindo $15 < \alpha < 20$.

Quando α , β e r_0 são constantes, as equações 3 e 1 são escritas com maior simplicidade como:

$$\frac{d}{dt} \ln S = -\frac{r_0}{N} \frac{dR}{dt} \quad (9)$$

e integrada como:

$$S(t) = S_0 \exp\left(-r_0 \frac{R - R_0}{N}\right) \quad (10)$$

em que S_0 e R_0 são constantes de integração a serem determinadas por $S_0 = S(t_0)$ e $R_0 = R(t_0)$.

Por 3, 9 pode ser reescrita como:

$$\gamma^{-1} \frac{dR}{dt} = N + R - S_0 \exp\left(-r_0 \frac{R - R_0}{N}\right) \quad (11)$$

que é uma equação separável, podendo ser resolvida, para o caso de coeficientes constantes,

por

$$\int_{R_0}^R \frac{du}{N + u - S_0 \exp(-r_0 \frac{u-R_0}{N})} = \gamma(t - t_0) \quad (12)$$

cuja primitiva não pode ser escrita a partir de funções elementares, tornando a equação pouco útil para as análises quantitativas feitas aqui.

2.4 Estimativa dos parâmetros

Conforme a epidemia estudada, há uma mudança nos valores do parâmetros, por isso é importante a definição e exposição dos valores utilizados nesse trabalho.

Já foram estimados, por diversas fontes, os valores dos parâmetros para a pandemia de COVID-19. Como colocado em [10], em [6] estima-se que $r_0 \approx 2.6$ com intervalo de confiança (1.5-3.5) e $\beta \approx 0.34$, baseando-se nos dados da pandemia na China. Esses valores implicam $\gamma \approx 0.13$. Já em [1], estima-se $\gamma \approx 0.145$ com intervalo de confiança (0.119-0.182), neste caso baseando-se nos dados da epidemia no Brasil. Ambos os dados implicam que uma pessoa passa em média $\gamma^{-1} \approx 7$ dias infectada podendo contaminar outras pessoas.

Nesse trabalho foi utilizado $0.119 < \gamma < 0.182$ e $\alpha = 15$, como estimado em [1] e [2]. A priori, foi calculado r_0 , a partir dos dados obtidos, da seguinte maneira:

$$r_0(t) = \frac{1}{1 - \mu} + \frac{\ddot{C}}{\gamma \dot{C}(1 - \mu)} \quad (13)$$

que pode ser obtido de 2 e 3 com μ dado por:

$$\mu(t) = \frac{\alpha}{\gamma N} (\gamma C + \dot{C}) \quad (14)$$

e para $\mu \ll 1$, longe da imunidade de rebanho, têm-se r_0 independente de α :

$$r_0(t) = 1 + \frac{\ddot{C}}{\gamma \dot{C}} \quad (15)$$

Enquanto a população estiver longe da imunidade de rebanho, a expansão e a extinção da epidemia não penderá nem de α nem de γ .

Porém, esta não é uma boa maneira de calculá-lo por depender da forma

derivada de C no tempo, pois é evidente que C tem muitos "ruídos" (que serão melhores tratados na próxima seção). Assim, ao derivar C no tempo, pode-se dizer que são tomados apenas os ruídos de C . Posteriormente, foi adotada outra maneira de estimar r_0 a partir de uma forma integral, não derivada.

A forma integral para estimar r_0 é encontrada integrando a equação 9

$$\ln \frac{S_0}{S(t)} = \frac{1}{N} \int_{t_0}^t r_0(u) \dot{R}(u) du \quad (16)$$

e então, supondo $R(t)$ monótona crescente, $\dot{R} > 0$, é possível mudar a variável de integração para obter:

$$\ln \frac{S_0}{S(t)} = \bar{r}_0 \frac{R(t) - R_0}{N} \quad (17)$$

em que \bar{r}_0 é o valor médio de r_0 no intervalo $[t_0, t]$ como no Teorema do Valor Médio para integrais. Formalmente a equação 17 é semelhante à 10.

Além disso, a função 14 pode ser utilizada para estimar α . De fato, seja t_* o instante em que $\ddot{C} = 0$, o ponto de inflexão da curva de casos acumulados, ou seja, o máximo da variação de \dot{C} , sabendo $r_0(t_*)$, têm-se:

$$\mu(t_*) = 1 - \frac{1}{r_0(t_*)} \quad (18)$$

Se forem conhecidos os valores de $C(t_*)$ e $\dot{C}(t_*)$, o valor de α é facilmente encontrado por 14.

3 Resultados

O resultado final do site pode ser encontrado em [11] e seu código fonte em [9]. No site, a análise estatística é feita para cada estado do Brasil, suas respectivas capitais, o país como um todo e algumas outras cidades do estado de São Paulo com a mostra de gráficos de vários tipos, alguns deles animados.

Em alguns casos, foi necessário o reajuste do parâmetro α por conta de uma diferença perceptível nos gráficos entre o valor estimado para o país e o valor real de certos estados. Já o valor de r_0 foi estimado para cada cidade, estado e para o país analisados estatisticamente, das duas maneiras apresentadas no texto, com um gráfico do valor diário para cada um.

4 Conclusão

O modelo SIR foi escolhido por sua simplicidade e efetividade como modelo epidemiológico. Partindo da premissa de cura com imunidade para todos os infectados, mesmo que essa não tenha sido ainda confirmada, pode-se dizer que esse tipo de modelo atendeu bem aos propósitos do trabalho que era uma maneira mais geral e técnica de análise de uma doença nova e pouco conhecida. Para infecções cuja cura não leva à imunidade, seria necessária a mudança para um modelo do tipo SIS [8], cujos infectados curados voltam a ser suscetíveis.

Referências

- [1] Saulo B. Bastos and Daniel O. Cajueiro. Modeling and forecasting the early evolution of the covid-19 pandemic in brazil. *arXiv*, June 16, 2020. Disponível em <<https://arxiv.org/abs/2003.14288>>.
- [2] Portal COVID-19 Brasil, 2020. Disponível em <<https://ciis.fmrp.usp.br/covid19/>>.
- [3] Davi Alves Monteiro Carvalho. Análise automática dos casos de covid-19: Arquivo html. Master's thesis, Universidade Estadual de Campinas, Campinas, 2021.
- [4] Ministério da Saúde, 2020. Disponível em <<https://covid.saude.gov.br/>>.
- [5] Jefferson Monção da Silva. Análise automática dos casos de covid-19: Análise estatística. Master's thesis, Universidade Estadual de Campinas, Campinas, 2021.
- [6] Chenlin Gu, Wei Jiang, Tianyuan Zhao, and Ban Zheng. Mathematical recommendations to fight against covid-19. *SSRN*, March 9, 2020. Disponível em <<https://ssrn.com/abstract=3551006> ou <http://dx.doi.org/10.2139/ssrn.3551006>>.
- [7] Mônica Helena Ribeiro Luiz. Modelos matemáticos em epidemiologia. Dissertação apresentada como requisito parcial para a obtenção do grau de mestre, Universidade Estadual Paulista "Júlio de Mesquita Filho", Rio Claro, 2012.
- [8] Lidiani Campos Auzani Pires. Estudo de modelos epidemiológicos determinísticos básicos que incluem transmissão vertical. Dissertação submetida como requisito parcial para a obtenção do grau de mestre em matemática aplicada, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2005.
- [9] A. Saa, 2020. Disponível em <<https://github.com/albertosaa/COVID>>.
- [10] A. Saa. Análise automática do painel coronavírus, 2020. Texto integralmente disponível em <<https://vigo.ime.unicamp.br/COVID/covid.pdf>>.
- [11] D. Carvalho; J. Monção; S. Zani; A. Saa, 2020. Painel Coronavirus, endereço eletrônico: <http://vigo.ime.unicamp.br/COVID/>.

- [12] A. Wassermann, 2013. Disponível em
<http://jsxgraph.org/wiki/index.php/Epidemiology:_The_SIR_model>.