Modelagem Epidemiológica do COVID-19

Modelo SIR/SEIR “0D” e “2D”

Prof. Marco Antonio Ridenti

Instituto Tecnológico de Aeronáutica
Departamento de Física

Relatório para comunicação das investigações teóricas voltadas à modelagem epidemiológica do COVID-19. Versão 2.0.

23 de março de 2020

Sumário

1 Objetivo

A presente versão deste relatório tem como objetivo a rápida comunicação da modelagem matemática para a descrição quantitativa da propagação da contaminação pelo COVID-19. Apresento dois tipos de modelo. O primeiro, mais simples, é um modelo global, pois considera a população como um todo e despreza a dinâmica local e a difusão territorial da epidemia. O segundo modelo resulta de um esforço para modelar a dinâmica espacial do vírus sobre o território de um país. A razão para considerar os dois casos está no fato de que o primeiro é de rápida implementação e permite captar as características gerais do problema. O segundo é mais realista, mas de implementação mais dispendiosa.

A abordagem "0-D" segue o padrão conhecido como SIR (susceptible-infected-recovered) e SEIR (susceptible-exposed-infected-recovered) (seção 2). Nós mostramos como essas abordagens podem ser extendidas para incluir a propagação regional das doenças infecciosas (seção 3).

2 Desenvolvimento Teórico - Modelo 0-D

2.1 Modelo SIR

No modelo SIR considera-se a evolução temporal da população susceptível (S), infectada (I) e recuperada (R). As equações diferenciais que regem a variação dessas populações é dada por

$$
\begin{aligned}
\frac{dS}{dt} &= \Lambda N - \mu_{eq} S - \beta \frac{S}{N} I && (1)\\
\frac{dI}{dt} &= \beta \frac{S}{N} I - (\mu + \gamma) I && (2)\\
\frac{dR}{dt} &= \gamma I - \mu_{eq} R && (3)\\
\frac{dN}{dt} &= \Lambda N - \mu_{eq} N - \mu_{cov} I && (4)\\
N &= S + I + R && (5)
\end{aligned}
$$

onde $\mu = \mu_{eq} + \mu_{cov}$ é a taxa de mortalidade total, composta da taxa de mortalidade de equilíbrio, μ_{eq}, mais a taxa de mortalidade da infecção pelo COVID-19, μ_{cov}. Os outros parâmetros são Λ, taxa de natalidade, β, taxa de contaminação e γ, taxa de recuperação. Note que a equação (4) resulta da soma das equações (1) a (3), não sendo portanto linearmente independente delas. Trata-se da equação para a população total em sua forma mais geral, admitindo aumento da população por conta da taxa de natalidade e diminuição por conta da taxa de mortalidade. As simplificações mais usuais habitualmente consideram uma situação de equilíbrio; no entanto, pelo conhecimento prévio sobre a doença, acredito ser necessário avaliar o impacto sobre a taxa de mortalidade.

2.2 Modelo SEIR

No modelo SEIR considera-se a evolução temporal da população susceptível (S), exposta(E), infectada (I) e recuperada (R). As equações diferenciais que regem a variação dessas populações é dada por

$$\frac{dS}{dt} = \Lambda - \mu_{eq}S - \beta\frac{S}{N}I \quad (6)$$

$$\frac{dE}{dt} = \beta\frac{S}{N}I - (\mu_{eq} + a)E \quad (7)$$

$$\frac{dI}{dt} = aE - (\mu + \gamma)I \quad (8)$$

$$\frac{dR}{dt} = \gamma I - \mu_{eq}R \quad (9)$$

$$\frac{dN}{dt} = \Lambda - \mu_{eq}N - \mu_{cov}I \quad (10)$$

$$N = S + E + I + R \quad (11)$$

$$(12)$$

onde $\mu = \mu_{eq} + \mu_{cov}$ é a taxa de mortalidade total, composta da taxa de mortalidade de equilíbrio, μ_{eq}, mais a taxa de mortalidade da infecção pelo COVID-19, μ_{cov}. Nesta equação foi introduzido o parâmetro a, isto é, a taxa de conversão de indivíduos expostos para infectados. Note que a equação (10) resulta da soma das equações (6) a (9), não sendo portanto linearmente independente delas. As mesmas considerações feitas anteriormente se aplicam a este caso.

3 Desenvolvimento Teórico - Propagação epidêmica sobre um território

A generalização para considerar a propação espacial (ou territorial) se baseia na equação da continuidade. Nesse caso, vamos trabalhar com a área total A_T de um território (digamos, o território brasileiro), e discretizá-lo em partes de igual área, A_i, que chamaremos de elemento de área. A forma das áreas é arbitrária, contanto que tenham sempre a mesma área, mas por simplificação podemos imaginá-la como sendo um quadrado. Cada elemento área A_i pode ser caracterizada pela coordenada geográfica central (latitude e longitude).

Em poucas palavras, o problema se resumirá a calcular a evolução das populações dentro dos elementos de área, levando-se em conta o fluxo da população para dentro e para fora dela. Antes disso, vamos introduzir alguma notação e definições importantes:

$$
\begin{aligned}
A_T &:= \text{`Área total do território.'} \\
A_i &:= \text{`}i\text{-ésimo elemento de área. Seu valor é simplesmente } A\text{.'} \\
A_T &= \sum_i A_i \\
s_i &= \frac{S}{A_i} := \text{`densidade de pessoas susceptíveis no } i\text{-ésimo elemento de área.'} \\
e_i &= \frac{E}{A_i} := \text{`densidade de pessoas expostas no } i\text{-ésimo elemento de área.'} \\
y_i &= \frac{I}{A_i} := \text{`densidade de pessoas infectadas no } i\text{-ésimo elemento de área.'} \\
r_i &= \frac{r}{A_i} := \text{`densidade de pessoas recuperadas no } i\text{-ésimo elemento de área.'} \\
n_i &= \frac{N}{A_i} := \text{`densidade total de pessoas no } i\text{-ésimo elemento de área.'}
\end{aligned}
$$

$$\Gamma_n(A_i) \;:=\; \text{‘fluxo total de indivíduos através do } i\text{-ésimo elemento de área.’}$$
$$\Gamma_s(A_i) \;:=\; \text{‘fluxo de indivíduos susceptíveis através do } i\text{-ésimo elemento de área.’}$$
$$\Gamma_e(A_i) \;:=\; \text{‘fluxo de indivíduos expostos através do } i\text{-ésimo elemento de área.’}$$
$$\Gamma_y(A_i) \;:=\; \text{‘fluxo de indivíduos infectados através do } i\text{-ésimo elemento de área.’}$$
$$\Gamma_r(A_i) \;:=\; \text{‘fluxo de indivíduos recuperados através do } i\text{-ésimo elemento de área.’}$$

Os fluxos tem tanto um contribuição positiva, isto é, de indivíduos que se deslocam de fora para dentro, $\Gamma^{\text{in}}(A_i)$, quanto uma contribuição negativa, isto é, de indivíduos que se deslocam de dentro para fora,$-\Gamma^{\text{out}}(A_i)$. Qualquer um dos fluxo acima definidos podem ser escritos como:

$$\Gamma_k(A_i) = \Gamma_k^{\text{in}}(A_i) - \Gamma_k^{\text{out}}(A_i)$$

onde k é índice que simplesmente representa uma das variáveis populacionais, isto é, $k = n, s, e, y, r$. Chamamos a atenção também para o fato de que o fluxo da população total é igual a soma dos fluxos de suas parcelas, isto é,

$$\Gamma_n(A_i) = \Gamma_s(A_i) + \Gamma_e(A_i) + \Gamma_y(A_i) + \Gamma_r(A_i)\ . \tag{13}$$

No modelo SIR basta omitir o segundo termo da soma. Vamos mostrar como esses fluxos podem ser modelados nas seções seguintes.

3.1 SIR - Espacial

As equações de taxa agora devem ser resolvidas para cada elemento de área, considerando-se os fluxos de saída e entrada.

$$\frac{ds_i}{dt} = \frac{\Gamma_s(A_i)}{A} + \Lambda n_i - \mu_{eq} s_i - \beta \frac{s_i}{n_i} y_i \tag{14}$$
$$\frac{dy_i}{dt} = \frac{\Gamma_y(A_i)}{A} + \beta \frac{s_i}{n_i} y_i - (\mu + \gamma) y_i \tag{15}$$
$$\frac{dr_i}{dt} = \frac{\Gamma_r(A_i)}{A} + \gamma y_i - \mu_{eq} r_i \tag{16}$$
$$\frac{dn_i}{dt} = \frac{\Gamma_n(A_i)}{A} + \Lambda n_i - \mu_{eq} n_i - \mu_{cov} y_i \tag{17}$$
$$n_i = s_i + y_i + r_i \tag{18}$$

Para obter a população de uma determinada parcela em todo território brasileiro, digamos, do número total de infectados, basta somar sobre todos os elementos de área e multiplicar o resultado pelo valor do elemento de área, *i.e.*

$$s = A \sum_i s_i \tag{19}$$

O acoplamento entre os diversas elementos de área é regido pelos fluxos de entrada, que como veremos mais a frente, é função dos fluxos de saída dos outros elementos de área.

3.2 SEIR - Espacial

As equações para o modelo SEIR podem ser escritas como

$$\begin{aligned}
\frac{ds_i}{dt} &= \frac{\Gamma_s(A_i)}{A} + \Lambda n_i - \mu_{eq} s_i - \beta \frac{s_i}{n_i} y_i && (20)\\
\frac{de_i}{dt} &= \frac{\Gamma_e(A_i)}{A} + \beta \frac{s_i}{n_i} y_i - (\mu_{eq} + a) e_i && (21)\\
\frac{dy_i}{dt} &= \frac{\Gamma_y(A_i)}{A} + a e_i - (\mu + \gamma) y_i && (22)\\
\frac{dr_i}{dt} &= \frac{\Gamma_r(A_i)}{A} + \gamma y_i - \mu_{eq} r_i && (23)\\
\frac{dn_i}{dt} &= \frac{\Gamma_n(A_i)}{A} + \Lambda n_i - \mu_{eq} n_i - \mu_{cov} y_i && (24)\\
n_i &= s_i + e_i + y_i + r_i && (25)
\end{aligned}$$

É importante considerar que os parâmetros do modelo podem agora ser considerados localmente, levando-se em conta as peculiaridades de cada região do território nacional.

3.3 Modelagem dos fluxos populacionais

O fluxo total de sáida em um elemento de área i pode ser prescrito, com base em conhecimento prévio sobre a região, ou por meio de algum tipo de modelo. É razoável considerar, por exemplo, que o fluxo de saída é proporcional à densidade populacional:

$$\Gamma_n^{\text{out}}(A_i) = \xi n_i \ . \tag{26}$$

Sabendo o fluxo total, podemos considerar que os fluxos das parcelas é proporcional às parcelas, isto é,

$$\begin{aligned}
\Gamma_s^{\text{out}}(A_i) &= \frac{s_i}{n_i}\Gamma_n^{\text{out}}(A_i) \\
\Gamma_e^{\text{out}}(A_i) &= \frac{e_i}{n_i}\Gamma_n^{\text{out}}(A_i) \\
\Gamma_y^{\text{out}}(A_i) &= \frac{y_i}{n_i}\Gamma_n^{\text{out}}(A_i) \\
\Gamma_r^{\text{out}}(A_i) &= \frac{r_i}{n_i}\Gamma_n^{\text{out}}(A_i)
\end{aligned}$$

Para modelar os fluxos de entrada em um determinado elemento de área A_i, nós temos que considerar que todas os outros elementos de área contribuem para o fluxo. Podemos escrever, de modo geral, o fluxo de entrada total em um elemento de área A_i como

$$\Gamma_n^{\text{in}}(A_i) = \sum_{j\neq i} f(d_{ij})\Gamma_n^{\text{out}}(A_j) \tag{27}$$

onde a soma é feita sobre todo os elementos de área, menos o elemento i, d_{ij} é a distância entre o elemento de área j e i e $f(d_{ij})$ é uma função decrescente de d_{ij}. Essa função deve ter a seguinte propriedade

$$\sum_{i\neq j} f(d_{ij}) = 1 \tag{28}$$

A função $f(d_{ij})$ deve ser interpretada como a fração do fluxo de população em A_j que se desloca para A_i. Os fluxos parciais também podem ser considerados proporcionais às parcelas populacionais, ou seja,

$$\begin{aligned}
\Gamma_s^{\text{in}}(A_i) &= \sum_{j\neq i} \frac{s_j}{n_j} f(d_{ij})\Gamma_n^{\text{out}}(A_j) \\
\Gamma_e^{\text{in}}(A_i) &= \sum_{j\neq i} \frac{e_j}{n_j} f(d_{ij})\Gamma_n^{\text{out}}(A_j) \\
\Gamma_y^{\text{in}}(A_i) &= \sum_{j\neq i} \frac{y_f}{n_j} f(d_{ij})\Gamma_n^{\text{out}}(A_j) \\
\Gamma_r^{\text{in}}(A_i) &= \sum_{j\neq i} \frac{r_j}{n_j} f(d_{ij})\Gamma_n^{\text{out}}(A_j)
\end{aligned}$$

3.4 Transformação para cidades/microrregiões

No modelo apresentado anteriormente, definimos um elemento de área arbitrário do território nacional. Note que as equações diferenciais desse modelo podem ser escritas independentemente do valor da área, bastando multiplicá-las membro a membro pela densidade A. Em geral, podemos trabalhar com o número total de municípios ou microrregiões em vez de trabalhar com elementos de área arbitrários. Tal abordagem facilita a implementação, a integração das informações do IBGE e a interpretação dos resultados.As equações, nessa abordagem, tomam a seguinte forma[1]:

$$
\begin{aligned}
\frac{dS_i}{dt} &= \Gamma_i^s + \Lambda N_i - \mu_{eq} S_i - \beta \frac{S_i}{N_i} I_i && (29) \\
\frac{dE_i}{dt} &= \Gamma_i^e + \beta \frac{S_i}{N_i} I_i - (\mu_{eq} + a) E_i && (30) \\
\frac{dI_i}{dt} &= \Gamma_i^y + a E_i - (\mu + \gamma) I_i && (31) \\
\frac{dR_i}{dt} &= \Gamma_i^r + \gamma I_i - \mu_{eq} R_i && (32) \\
\frac{dN_i}{dt} &= \Gamma_i^n + \Lambda N_i - \mu_{eq} N_i - \mu_{cov} I_i && (33) \\
N_i &= S_i + E_i + Y_i + R_i && (34)
\end{aligned}
$$

Agora cada população ou parcela populacional (S_i, E_i, I_i, R_i) é indexada pela i-ésima micorregião ou i-ésimo munícipio. A interpretação para os fluxos e para a função $f(d_{i,j})$ é análoga ao caso anterior. Note também que mudamos a notação do fluxo; o sub-índice i indica a microrregião/município e o índice superescrito indica a parcela populacional.

3.5 Inclusão de faixa etária

A inclusão da faixa etária pode ser realizada a partir da quebra da população em estratos de faixa etária. Dessa forma, uma dada variável de população, digamos, a população total $N_{i,k}$, é agora indexada pela sua i-ésima microrregião e sua k-ésima faixa etária. No Brasil, por exemplo, temos $M = 558$ microrregiões e podemos,

[1] A partir desse ponto, vamos escrever as equações sempre no formalismo SEIR.

por exemplo, trabalhar com faixas etárias espaçadas em cinco anos até 80 ou mais, totalizando $Q = 17$ faixas etárias. Considerando que temos quatro parcelas populacionais (S, E, I, R), teremos no total $4 \times 558 \times 17 = 37944$ variáveis dependentes do tempo e $\times 558 \times 17 = 9486$ sistemas de equações diferenciais ordinárias a resolver. Apesar de parecer muito, é um cálculo factível para os padrões computacionais atuais. Antes de escrever as equações, é necessário introduzir uma nova notação.

$$\begin{aligned}
S_{ik} &:= \text{‘número de indivíduos susceptíveis na } i\text{-ésima região e } k\text{-ésima faixa etária.’} \\
E_{ik} &:= \text{‘número de indivíduos expostos na } i\text{-ésima região e } k\text{-ésima faixa etária.’} \\
I_{ik} &:= \text{‘número de indivíduos infectados na } i\text{-ésima região e } k\text{-ésima faixa etária.’} \\
R_{ik} &:= \text{‘número de indivíduos recuperados na } i\text{-ésima região e } k\text{-ésima faixa etária.’} \\
N_{ik} &:= \text{‘número total de indivíduos na } i\text{-ésima região e } k\text{-ésima faixa etária.’}
\end{aligned}$$

$$\begin{aligned}
\Gamma^n_{ik} &:= \text{‘fluxo total de indivíduos através da } i\text{-ésima região, na } k\text{-ésima faixa etária.’} \\
\Gamma^s_{ik} &:= \text{‘fluxo total de susceptíveis através da } i\text{-ésima região, na } k\text{-ésima faixa etária.’} \\
\Gamma^e_{ik} &:= \text{‘fluxo total de expostos através da } i\text{-ésima região, na } k\text{-ésima faixa etária.’} \\
\Gamma^y_{ik} &:= \text{‘fluxo total de infectados através da } i\text{-ésima região, na } k\text{-ésima faixa etária.’} \\
\Gamma^r_{ik} &:= \text{‘fluxo total de recuperados através da } i\text{-ésima região, na } k\text{-ésima faixa etária.’}
\end{aligned}$$

$$\begin{aligned}
S_i &= \sum_k S_{ik} \quad ; \quad \Gamma^s_i = \sum_k \Gamma^s_{ik} \\
E_i &= \sum_k E_{ik} \quad ; \quad \Gamma^e_i = \sum_k \Gamma^e_{ik} \\
I_i &= \sum_k I_{ik} \quad ; \quad \Gamma^y_i = \sum_k \Gamma^y_{ik} \\
R_i &= \sum_k R_{ik} \quad ; \quad \Gamma^r_i = \sum_k \Gamma^r_{ik} \\
N_i &= \sum_k N_{ik} \quad ; \quad \Gamma^n_i = \sum_k \Gamma^n_{ik}
\end{aligned}$$

As equações podem então ser escritas como:

$$\frac{d}{dt}S_{ik} = \Gamma^s_{ik} + \delta_{1k}\Lambda_{ik}N_{ik} - \mu_{eq,ik}S_{ik} - \sum_j \beta_{i,kj}\frac{S_{ik}}{N_i}I_{ij} \tag{35}$$

$$\frac{d}{dt}E_{ik} = \Gamma^e_{ik} + \sum_j \beta_{i,kj}\frac{S_{ik}}{N_i}I_{ij} - (\mu_{eq,ik} + a_{ik})E_{ik} \tag{36}$$

$$\frac{d}{dt}I_{ik} = \Gamma^y_{ik} + a_{ik}E_{ik} - (\mu_{ik} + \gamma_{ik})I_{ik} \tag{37}$$

$$\frac{d}{dt}R_{ik} = \Gamma^r_{ik} + \gamma_{ik}I_{ik} - \mu_{eq,ik}R_{ik} \tag{38}$$

$$\frac{d}{dt}N_{ik} = \Gamma^n_{ik} + \delta_{1k}\Lambda_{ik}N_{ik} - \mu_{eq,ik}N_{ik} - \mu_{cov,ik}I_{ik} \tag{39}$$

$$N_{ik} = S_{ik} + E_{ik} + Y_{ik} + R_{ik} \tag{40}$$

onde δ_{1k} é a função delta de Kronecker (1 se $k = j$ e 0 noutro caso) e $\beta_{i,kj}$ é a taxa de exposição que um indivíduo susceptível da faixa etária k tem a um indivíduo infectado da faixa etária j, na região i. Note na notação aqui adotada a dependência dos parâmetros com a região e faixa etária estão explicitadas por meio dos índices subscritos, por exemplo, $\mu_{eq,ik}$ é a taxa de mortalidade na faixa etária k na região i.

Em uma escala temporal maior do que a escala típica das epidemias, o envelhecimento da população produz um fluxo entre as faixas etárias, apenas positivo. Nas equaçõe acima, não levamos em consideração esse fluxo contínuo, desprezível contanto que as faixas etárias não sejam mais estreitas do que um ano.

3.5.1 Modelo "0D" com faixas etárias

É possível escrever as equações SEIR considerando as faixas etárias, desconsiderando no entanto sua distribuição espacial. Isso pode ser útil em modelos globais em que a informação sobre as taxas de infecção por faixa etária é importante. Nesse caso, os índices subscritos indicam apenas a faixa etária, e não mais a região.

$$
\begin{aligned}
\frac{d}{dt}S_k &= \Gamma_k^s + \delta_{1k}\Lambda_k N_k - \mu_{eq,k}S_k - \sum_j \beta_{kj}\frac{S_k}{N}I_j & (41)\\
\frac{d}{dt}E_k &= \Gamma_k^e + \sum_j \beta_{kj}\frac{S_k}{N}I_j - (\mu_{eq,k} + a_k)E_k & (42)\\
\frac{d}{dt}I_k &= \Gamma_k^y + a_k E_k - (\mu_k + \gamma_k)I_k & (43)\\
\frac{d}{dt}R_k &= \Gamma_k^r + \gamma_k I_k - \mu_{eq,k}R_k & (44)\\
\frac{d}{dt}N_k &= \Gamma_k^n + \delta_{1k}\Lambda_k N_k - \mu_{eq,ik}N_k - \mu_{cov,k}I_k & (45)\\
N_k &= S_k + E_k + Y_k + R_k & (46)
\end{aligned}
$$