

Nota Técnica SEIR-Bayes

Daniel Severo
3778 Healthcare
São Paulo, Brazil
severo@3778.care

Giuliano Netto Flores Cruz
Florianopolis, Brazil
cruz.gnf@gmail.com

I. MODELO SEIR-BAYES

Modelos compartimentados são uma família de técnicas utilizadas para estudar e prever a dinâmica temporal de contagem e fluxo entre grupos de objetos de estudo [5]. No caso de doenças infecciosas, são utilizados para prever o alcance e impacto da doença. A população de estudo é particionada em compartimentos disjuntos de forma que os indivíduos em um mesmo compartimento possuam características homogêneas. O saldo total de pessoas em cada grupo é dependente do tempo, pois existe um fluxo de entrada e saída em cada compartimento. A dinâmica do fluxo define o comportamento transiente e estacionário do modelo, e sua intensidade é variável com o tempo; podendo ser desde uma constante até uma função do número total de pessoas por compartimento. Este modelo permite projetar a curva epidemiológica para um dado grupo populacional considerando os parâmetros devidos.

Neste trabalho, foi considerado um modelo com 4 compartimentos: **Suscetíveis** (população não infectada, porém passível de infecção), **Expostas** (infectadas, porém não-contagiosas), **Infectadas** (infectadas e contagiosas), e **Removidas** (removidas do estudo por óbito ou imunidade); denominado SEIR [5].

A evolução temporal (variável t) é discretizada a dias. Diariamente, existe uma probabilidade de um indivíduo migrar de compartimento (Tabela I). Este valor é variável e depende da contagem de pessoas nos compartimentos S_t , E_t , I_t , e R_t ; assim como os parâmetros α^{-1} (período médio de incubação), γ^{-1} (período médio infeccioso) e r_0 (número básico de reprodução).

TABLE I
EXPRESSÕES PARA ESTIMAÇÃO DOS PARÂMETROS DO MODELO SEIR

De	Para	Probabilidade
Suscetíveis	Expostas	$p_{SE}(t) = 1 - \exp(\frac{-r_0 \gamma I_t}{N})$
Expostas	Infectadas	$p_{EI}(t) = 1 - \exp(-\alpha)$
Infectadas	Removidas	$p_{IR}(t) = 1 - \exp(-\gamma)$

Esta componente estocástica resulta em um comportamento não-determinístico para a contagem e fluxo dos compartimentos. Note que $1 - \exp(-x)$ é próximo de x para valores de x próximos a zero. Logo, para períodos de incubação e infecção elevados (acima de 1 dia), temos $p_{EI} = \alpha$ e $p_{IR} = \gamma$; que representam as taxas de infecção (razão entre infectados e expostos) e remoção (razão entre removidos e infectados), respectivamente.

Para capturar as incertezas presentes na literatura para os parâmetros α e γ , foi utilizada uma distribuição LogNormal com 95% da densidade dentro dos intervalos encontrados na literatura (Tabela II). Já o r_0 foi estimado a partir da série histórica de incidência de casos, cujo detalhamento está na seção a seguir. As condições iniciais do modelo $S(0)$, $E(0)$, $I(0)$ e $R(0)$ são consideradas como dadas.

TABLE II
VALORES DE REFERÊNCIA PARA OS PARÂMETROS X E Y

Parâmetro	Limite Inferior	Limite Superior	Referência
α^{-1}	4,1 dias	7 dias	[6], [10]
γ^{-1}	10 dias	14 dias	[1]

Formalmente, para os fluxos entre compartimentos, temos

$$SE_{t+1} \sim \text{Binomial}(S_t, p_{SE}(t) | r_0, \gamma) \quad (1)$$

$$EI_{t+1} \sim \text{Binomial}(E_t, p_{EI}(t) | \alpha) \quad (2)$$

$$IR_{t+1} \sim \text{Binomial}(I_t, p_{IR}(t) | \gamma) \quad (3)$$

A partir dos fluxos é possível calcular a contagem

$$S_{t+1} = -SE_t \quad (4)$$

$$E_{t+1} = SE_t - EI_t \quad (5)$$

$$I_{t+1} = EI_t - IR_t \quad (6)$$

$$R_{t+1} = IR_t \quad (7)$$

II. ESTIMATIVA DO NÚMERO DE REPRODUÇÃO

O número de reprodução foi estimado com base na metodologia descrita em [3], [9]. Assume-se que a incidência observada a cada tempo t , I_t , seja descrita adequadamente por uma distribuição de Poisson cujo parâmetro λ_t se traduz em uma função do número de reprodução instantâneo, r_t , e do potencial infeccioso total, ao longo de todos os indivíduos infectados no tempo t , Λ_t . Formalmente:

$$I_t \mid r_t, \Lambda_t \sim \text{Poisson}(\lambda_t) \quad (8)$$

$$\lambda_t = r_t \cdot \Lambda_t \quad (9)$$

Na prática, considera-se r_t constante ao longo de uma janela temporal de tamanho τ , de forma que a equação 9 torna-se $\lambda_t = r_t \cdot \Lambda_{t-\tau}$. O parâmetro $\Lambda_{t-\tau}$ pode então ser

interpretado como o potencial infeccioso total ao longo do período $[t - \tau, t]$, sendo portanto uma função da incidência em cada passo temporal prévio a I_t , denotada I_{t-s} para $s \in [1, t - \tau]$, ponderada pela probabilidade de infecção secundária em cada período de tempo s , denotada w_s :

$$\Lambda_{t-\tau} = \sum_{s=1}^{t-\tau} I_{t-s} w_s \quad (10)$$

O parâmetro w_s é uma função de massa de probabilidade computada a partir da distribuição estimada para o intervalo serial da doença, isto é, o tempo entre o aparecimento de sintomas em uma pessoa infectada (infectior) e o aparecimento de sintomas em pessoas por ela infectadas (infectees). O intervalo serial é comumente usado como proxy para o tempo entre uma infecção primária e as infecções secundárias dela resultantes. Aqui, foi assumido que o intervalo serial é descrito adequadamente por uma distribuição Gamma com média 4,89 e desvio padrão 1,48. Tal parametrização garante 95% de densidade no intervalo $[2, 4, 8, 3]$ e foi escolhida com base nas recentes estimativas de intervalo serial para SARS-Cov-2 [4], [8], [11], [12].

A inferência sobre r_t se dá via atualização Bayesiana. Sua distribuição a posteriori, condicional aos dados de incidência entre tempo 0 e tempo t e à função de probabilidade w_s , é computada a partir da verossimilhança da incidência observada no intervalo $[t - \tau, t]$ e da distribuição a priori assumida para r_t :

$$P(r_t | I_{[0,t]}, w_s) \propto P(I_{[t-\tau,t]} | I_{[0,t-\tau-1]}, w_s) \cdot P(r_t) \quad (11)$$

Assumindo a priori que r_t segue uma distribuição Gamma, pode-se chegar a uma fórmula fechada para r_t a posteriori haja vista o fato de ser Gamma uma distribuição conjugada à de Poisson, utilizada para computar a verossimilhança. Sendo assim, $P(r_t | I_{[0,t]}, w_s)$ é descrita adequadamente pela distribuição $\Gamma(k, \theta)$ em que k e θ são parâmetros de forma e escala, respectivamente, dados por:

$$k = k_0 + \sum_{j=t-\tau}^t I_j \quad (12)$$

$$\theta = \left(\sum_{j=t-\tau}^t \Lambda_j + \frac{1}{\theta_0} \right) \quad (13)$$

em que k_0 e θ_0 são os parâmetros correspondentes da distribuição a priori de r_t .

Com base em estimativas anteriores do número básico de reprodução, assumiu-se $k_0 = 5,12$ e $\theta_0 = 0,64$, correspondendo a uma média e um desvio padrão de aproximadamente 3,3 e 1,45, respectivamente, e posicionando 95% da densidade a priori entre 1,1 e 6,7. Em linhas gerais, essa parametrização abrange as estimativas de 14 estudos anteriores revisados por [7]. Foi utilizada a janela temporal de $\tau = 12$ dias uma vez que janelas menores geraram estimativas instáveis, potencialmente sensíveis a eventual represamento de confirmação de

casos diários - em função de falta de testes diagnósticos ou capacidade de processamento destes. Janelas maiores foram descartadas por impossibilitar estimativa para maioria dos estados brasileiros nos quais não haveria dados de incidência suficientes. Estados com dados insuficientes tiveram seu r_t imputado com o valor correspondente ao r_t do Brasil como um todo. O valor de r_t mais recente disponível para cada caso foi utilizado como entrada no modelo SEIR. A metodologia aqui descrita foi implementada em python 3 e reproduz os resultados da função “estimate_R()”, do pacote de R EpiEstim, utilizando o método “parametric_si” [2].

REFERENCES

- [1] Jun Chen, Tangkai Qi, Li Liu, Yun Ling, Zhiping Qian, Tao Li, Feng Li, Qingnian Xu, Yuyi Zhang, Shuibao Xu, et al. Clinical progression of patients with covid-19 in shanghai, china. *Journal of Infection*, 2020.
- [2] Anne Cori. Epiestim: Estimate time varying reproduction numbers from epidemic curves (version 2.2.1). <https://CRAN.R-project.org/package=EpiEstim>, 2019.
- [3] anne cori, neil m ferguson, christophe fraser, and simon cauchemez. a new framework and software to estimate time-varying reproduction numbers during epidemics. *american journal of epidemiology*, 178(9):1505–1512, 2013.
- [4] Zhanwei Du, Xiaoke Xu, Ye Wu, Lin Wang, Benjamin J Cowling, and Lauren Ancel Meyers. The serial interval of covid-19 from publicly reported confirmed cases. *medRxiv*, 2020.
- [5] William Ogilvy Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927.
- [6] Qun Li, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou, Yeqing Tong, Ruiqi Ren, Kathy SM Leung, Eric HY Lau, Jessica Y Wong, et al. Early transmission dynamics in wuhan, china, of novel coronavirus-infected pneumonia. *New England Journal of Medicine*, 2020.
- [7] Ying Liu, Albert A Gayle, Annelies Wilder-Smith, and Joacim Rocklöv. The reproductive number of covid-19 is higher compared to sars coronavirus. *Journal of travel medicine*, 2020.
- [8] Hiroshi Nishiura, Natalie M Linton, and Andrei R. Akhmetzhanov. Serial interval of novel coronavirus (2019-ncov) infections. *medRxiv*, 2020.
- [9] rn thompson, je stockwin, rd van gaalen, ja polonsky, zn kamvar, pa demarsh, e dahlqwist, s li, eve miguel, t jombart, et al. improved inference of time-varying reproduction numbers during infectious disease outbreaks. *epidemics*, 29:100356, 2019.
- [10] WHO. Report of the who-china joint mission on coronavirus disease 2019 (covid-19)., 2020.
- [11] Joseph T Wu, Kathy Leung, and Gabriel M Leung. Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study. *The Lancet*, 395(10225):689–697, 2020.
- [12] Shi Zhao, Daozhou Gao, Zian Zhuang, Marc Chong, Yongli Cai, Jinjun Ran, Peihua Cao, Kai Wang, Yijun Lou, Weiming Wang, et al. Estimating the serial interval of the novel coronavirus disease (covid-19): A statistical analysis using the public data in hong kong from january 16 to february 15, 2020. *medRxiv*, 2020.