

Até que a morte os separe: percepções de um modelo matemático básico na pandemia de Covid-19 em andamento

Moacyr Alvim Silva

Escola de Matemática Aplicada, Fundação Getúlio Vargas, Rio de Janeiro

Helio Schechtman

Programa de Computação Científica, Fundação Oswaldo Cruz, Rio de Janeiro

Max Souza

Instituto de Matemática e Estatística, Universidade Federal Fluminense, Niterói

Esta versão: 4 de Setembro de 2020

Introdução

A evolução de epidemias de doenças infecciosas é monitorada pela série temporal de casos, ou óbitos, conforme pode ser visualizado na figura 1, extraído do website Our World in Data [1]. Nota-se que existe uma região aproximadamente linear logo no início (marcada pela linha vermelha). Essa região linear no gráfico com escala logarítmica no eixo vertical exhibe o conhecido fenômeno do crescimento exponencial do número de óbitos no início de epidemias. Embora a evolução temporal da epidemia em seu início seja bem descrita por uma função exponencial, é bastante temerário avaliar e estimar o comportamento da epidemia quando de posse de somente destes dados iniciais. Neste boletim pretendemos investigar as seguintes perguntas: quais informações podemos extrair dos dados de óbitos no início do surto epidêmico e quais informações permanecem ocultas, mesmo supondo dados confiáveis.

Para isso, utilizamos um modelo compartimental **SICRD** — Suscetível (S), Infeccioso (I), Criticamente infeccioso (C), Recuperado (R) e Morto (D) — para capturar a dinâmica inicial de doenças como o COVID-19 [2]. A escolha deste modelo se deve ao fato da mortalidade aparecer explicitamente como um parâmetro do modelo, e não combinado outros parâmetros, como é comum em muitos modelos similares. Essa característica facilita a análise, já que partimos da suposição de que os únicos dados observados (razoavelmente confiáveis) referem-se ao número de mortes pela doença. A figura 2 apresenta uma representação esquemática dos diversos compartimentos, ou estágios, pelos quais um indivíduo acometido da doença pode passar, segundo este modelo. Ainda que bastante simples é, como outros modelos mais complexos, não-linear devido ao termo que descreve a conexão entre os compartimentos Suscetível (S), Infeccioso (I) e Criticamente Infeccioso (C). Embora seja não-linear, é possível representá-lo por um modelo linear em uma região vizinha ao DFE (Disease Free Equilibrium - em português Equilíbrio Livre de Doença), ou seja quando a população está começando a se infectar com a doença.

A calibração deste modelo linear pode ser efetuada com a utilização dos dados de óbitos da doença em sua fase linear. A utilização dos dados de óbitos, ao invés do número de casos (pessoas infectadas) minimiza a incerteza devido à sub-notificação de casos e da possível existência de portadores assintomáticos da doença. Este modelo linear então calibrado gerará valores e/ou faixas de valores para os diversos parâmetros do modelo compartimental original não-linear. Os dados de óbitos obtidos pelo modelo podem então ser comparados aos dados reais. A partir desta comparação pode-se fazer inferências seja da evolução da epidemia seja sobre a eficácia das medidas implementadas para mitigação da epidemia.

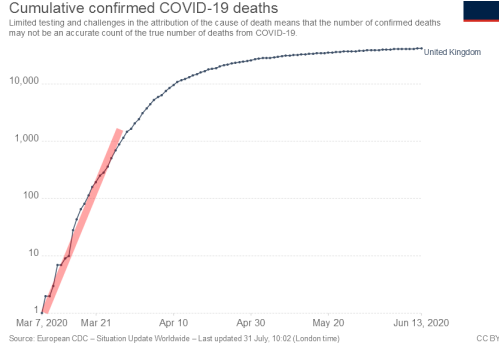


Figura 1: Óbitos acumulados de covid-19 no início da epidemia no Reino Unido [1]. Note a região aproximadamente linear logo no início marcada pela linha vermelha.

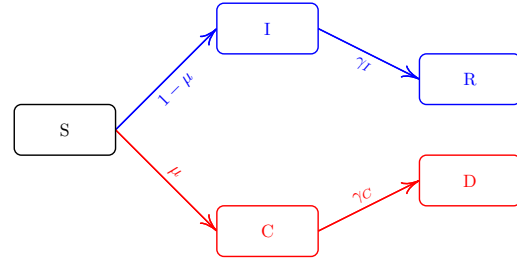


Figura 2: Representação esquemática do modelo SICRD.

Materiais e métodos

Modelo matemático

O modelo compartimental **SICRD** é dado pelas seguintes equações:

$$\begin{aligned}
 \dot{S} &= -\frac{\beta}{N}S[I + \delta C], \\
 \dot{I} &= (1 - \mu)\frac{\beta}{N}S[I + \delta C] - \gamma_I I, \\
 \dot{C} &= \mu\frac{\beta}{N}S[I + \delta C] - \gamma_C C, \\
 \dot{R} &= \gamma_I I, \\
 \dot{D} &= \gamma_C C.
 \end{aligned} \tag{1}$$

Para este modelo, utilizando-se a técnica de matriz de próxima geração, obtém-se o número de reprodução básica

$$\mathcal{R}_0 = (1 - \mu)\mathcal{R}_0^I + \mu\mathcal{R}_0^C. \tag{2}$$

onde

$$\mathcal{R}_0^I = \frac{\beta}{\gamma_I} \text{ e } \mathcal{R}_0^C = \frac{\beta\delta}{\gamma_C}$$

Considerando que no início da epidemia $S \approx N$ gera-se o modelo linearizado próximo ao ponto do sistema onde não há doentes, ou DFE - *disease free equilibrium*. A evolução temporal dos números de casos nos compartimentos **C** e **I** é dada então pela equação

$$\dot{Y} = AY$$

onde

$$A = \begin{pmatrix} (1 - \mu)\beta - \gamma_I & (1 - \mu)\beta\delta \\ \beta\mu & \beta\delta\mu - \gamma_C \end{pmatrix} \text{ e } Y = \begin{pmatrix} I \\ C \end{pmatrix}$$

sendo a solução dada por

$$Y(t) = Y_0 \exp(tA).$$

O número de óbitos (compartimento **D**) para o modelo proposto é obtido pela equação

$$\dot{D} = \gamma_C C.$$

Substituindo-se C pela expressão obtida pelo modelo linear acima obtém-se o número de mortes $D(t)$ no tempo t no início da epidemia, quando a dinâmica é preponderantemente linear

$$D(t) = D_0 + \frac{K_1}{\lambda_1} (\exp(\lambda_1 t) - 1) + \frac{K_2}{\lambda_2} (\exp(\lambda_2 t) - 1).$$

Por outro lado, temos que

$$\lambda_1 \lambda_2 = \det(A) = \gamma_I \gamma_C (1 - \mathcal{R}_0). \quad (3)$$

Considerando que o modelo refere-se ao caso de doença epidêmica, $\mathcal{R}_0 > 1$, $\gamma_I > 0$ e $\gamma_C > 0$, então não é difícil de ver que ambos os autovalores λ_1 e λ_2 da matriz A serão reais e possuirão sinais opostos. No que se segue, suporemos, sem perda de generalidade, que $\lambda_1 > 0$ e $\lambda_2 < 0$.

Neste caso, pode-se verificar que as constantes K_1 e K_2 são as seguintes:

$$K_1 = \frac{I_0 + (d - \lambda_1)C_0}{\lambda_1 - \lambda_2} \quad \text{e} \quad K_2 = \frac{I_0 + (d - \lambda_2)C_0}{\lambda_1 - \lambda_2}$$

sendo C_0 e I_0 as quantidades iniciais de indivíduos Críticos e Infectados não-críticos, respectivamente, e $d = \beta\delta\mu - \gamma_C$.

Ajuste dos dados

O procedimento de ajuste foi dividido em um processo de dois estágios: *o ajuste linear* e *o bracketing*.

Ajuste linear

A calibração deste modelo em sua fase linear foi efetuada através do método de mínimos quadrados com a utilização dos dados de óbito de COVID-19. Apesar de estarem disponíveis métodos mais sofisticados e flexíveis para a calibração de curvas exponenciais do que o método dos mínimos quadrados [3], para os nossos propósitos este método se mostrou suficiente. Para evitar ruídos e comportamento estocástico significativo, o modelo foi ajustado aos dados a partir do momento em que o total de óbitos ultrapassou 50 indivíduos. O ajuste aos dados nesta fase linear foi realizado em uma série de janelas discretas de 7 a 28 dias, sendo escolhida a que melhor ajuste apresentou.

Curiosamente, esta calibração pode ser feita sem o conhecimento dos verdadeiros valores dos parâmetros do modelo e permite um ajuste **universal**. O ajuste aos dados na fase linear não é significativamente sensível à escolha de K_2 , uma vez que o ajuste é efetuado apenas no período onde o número de óbitos é maior do que 50. Por isso, o tempo decorrido entre a introdução da doença na população e as 50 mortes já é grande o suficiente para que a exponencial referente a K_2 na expressão do número de óbitos seja muito pequena quando comparada a exponencial referente a K_1 . A constante K_2 só teria um papel relevante se fosse grande quando comparada a K_1 . Mas é razoável supor exatamente o contrário, que K_1 tenha um valor bem superior ao de K_2 . Isto se justifica pois os indivíduos (I)nfectados e (C)ríticos introduzidos na população foram provavelmente importados de uma localidade onde a epidemia já estava em estágio mais avançado, ou seja, a proporção inicial entre Infectados e Críticos deve ser próxima dessa proporção vinda do autovetor correspondente a λ_1 . Quando essa proporção inicial é próxima a deste autovetor temos $K_2 \approx 0$.

Relembrando as equações (2) e (3) temos

$$\mathcal{R}_0 = 1 - \frac{\lambda_1 \lambda_2}{\gamma_I \gamma_C}$$

E, assumindo que $\gamma_I > \gamma_C$ (suposição razoável para Covid-19 [4]) e como necessariamente deve-se ter $\frac{\beta}{\gamma_I} = (1 - \mu)\mathcal{R}_0^I > 0$ e $\frac{\beta\delta}{\gamma_C} = \mu\mathcal{R}_0^C > 0$, obtém-se consequentemente que

$$-\gamma_I < \lambda_2 < -\gamma_C$$

Assim, a única restrição imposta ao autovalor λ_2 é que esteja contido no intervalo acima. Seu valor é, de outra forma, livre. Para todas as opções dos demais parâmetros livres (I_0 , β , μ , δ , γ_I e γ_C) a dinâmica linear é indistinguível - embora a dinâmica não-linear seja diferente.

Este procedimento de ajuste aos dados na fase linear produz então os parâmetros K_1 e λ_1 do modelo.

Bracketing

O procedimento de ajuste aos dados na fase linear fixa os parâmetros K_1 e λ_1 do modelo. Entretanto, a dinâmica não-linear dependerá dos demais parâmetros, mormente λ_2 e μ (ou δ). O ajuste aos dados nesta etapa também foi realizada por mínimos quadrados. A dinâmica de longo prazo é controlada principalmente pelo valor de μ — as diferenças na dinâmica não-linear dentro da faixa de valores possíveis de \mathcal{R}_0 não se mostraram significativas.

O período de tempo usado para realizar o "bracketing", ou seja obter os super/sub-ajustes foi de duas semanas após a janela de tempo usada para a calibração do modelo linear.

Lembrando que $\mathcal{R}_0 = 1 - \frac{\lambda_1 \lambda_2}{\gamma_I \gamma_C}$ e $-\gamma_I < \lambda_2 < -\gamma_C$, conseqüentemente os valores possíveis de R_0 estarão compreendidos entre dois extremos

$$\underline{\mathcal{R}_0} = 1 + \frac{\lambda_1}{\gamma_I} < \mathcal{R}_0 < 1 + \frac{\lambda_1}{\gamma_C} = \overline{\mathcal{R}_0} \quad (4)$$

O ajuste do parâmetro μ , a partir dos dados da dinâmica não-linear e da adequação de R_0 a faixa de valores permitida é realizada da seguinte forma:

Super-ajuste O menor valor de μ para o maior \mathcal{R}_0 admissível que permite que a solução fique acima ou igual aos dados;

Sub-ajuste O maior valor de μ para o menor \mathcal{R}_0 admissível, que permite que a solução permaneça abaixo ou igual aos dados.

Resultados

Os resultados abaixo foram obtidos utilizando-se código em Python simples e disponível em anexo a este boletim. Ressalta-se que

- O encaixe linear é universal;
- Para o "bracketing" foi usado $\gamma_I = \frac{1}{7}$ e $\gamma_C = \frac{1}{21}$ valores razoáveis para a modelagem de Covid-19 obtidos de [4].
- O ajuste da fase linear considerou os dados após a notificação de 50 óbitos. O período anterior a isto está presente nos gráficos, apesar de não terem sido considerados para o ajuste, e estão rotulados como *out of sample*

Espanha

- $\lambda_1 = 0,22117434$ e $K_1 = 0,15445339$
- Coeficiente de determinação $R^2 = 0,9850592304071338$
- Período de ajuste linear: 25 – 38 (14 dias)

	\mathcal{R}_0	I_0	δ	μ
Super-ajuste	6,457	132939	120336	6×10^{-3}
Sub-ajuste	2,819	920	1,15	7×10^{-4}

Tabela 1: Espanha - parâmetros de ajuste.

Ajuste linear e "bracketing"

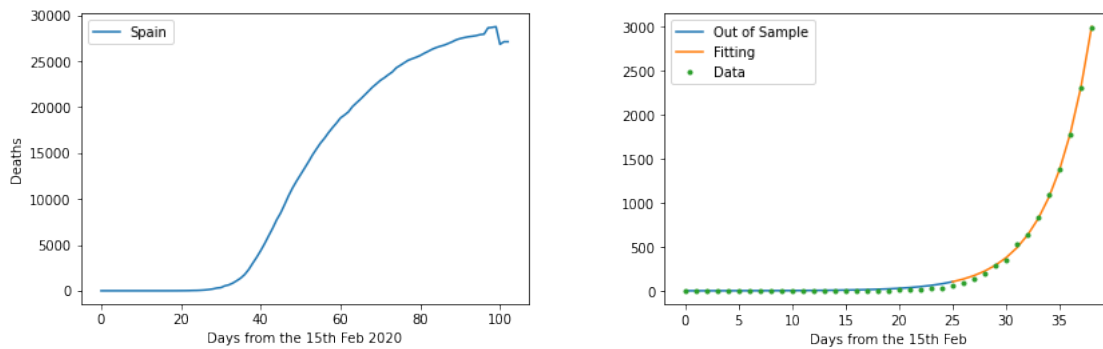


Figura 3: Espanha - dados epidemiológicos e curva ajustada.

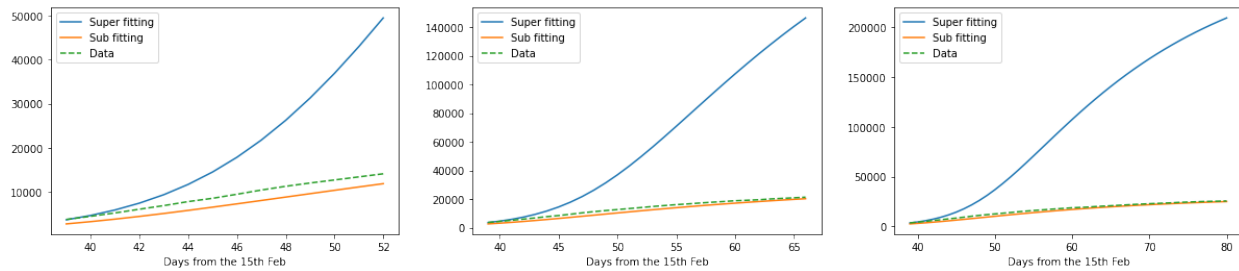


Figura 4: Espanha - Super / Sub-ajuste após o ajuste linear: 14 dias (esquerda); 28 dias (centro); 42 dias (direita).

Perspectiva de longo prazo

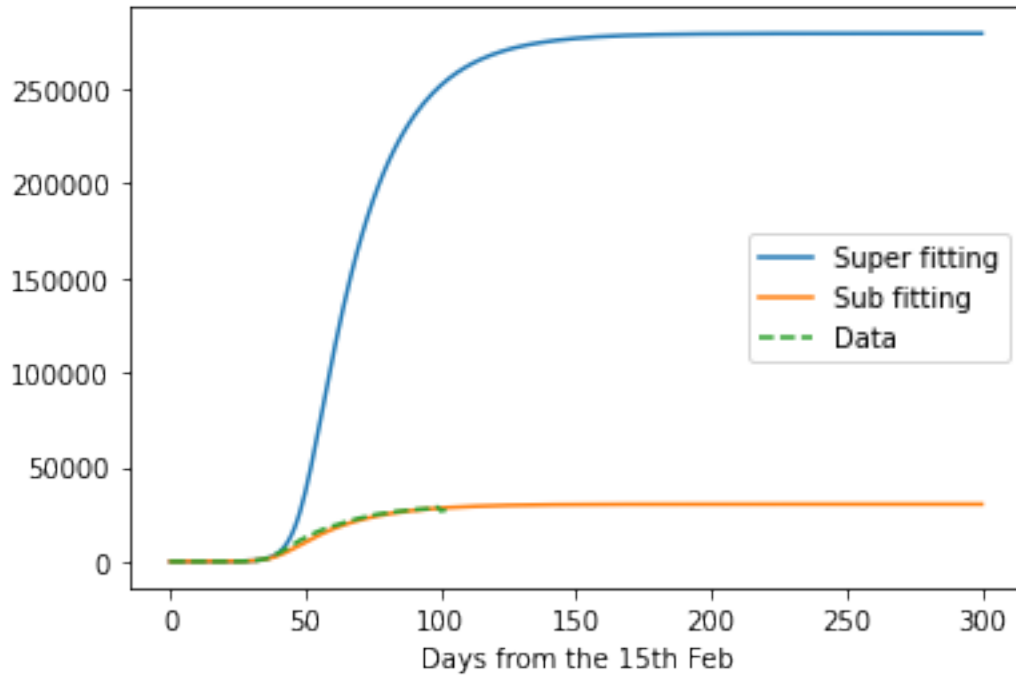


Figura 5: Espanha - A evolução "completa" das curvas de bracketing, juntamente com os dados disponíveis até o momento.

Suécia

- $\lambda_1 = 0,20777427$ e $K_1 = 0,01681574$
- Coeficiente de determinação $R^2 = 0,8488406611476258$
- Período de ajuste linear: 39 – 49 (11 dias)

	\mathcal{R}_0	I_0	δ	μ
Super-ajuste	5,36	69788	692274	1×10^{-3}
Sub-ajuste	2,45	178	2,55	3×10^{-4}

Tabela 2: Suécia - parâmetros de ajuste.

Ajuste linear e "bracketing"

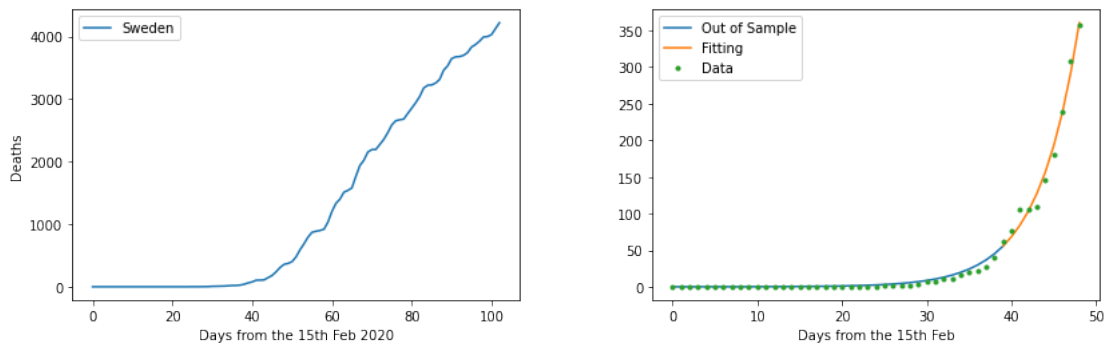


Figura 6: Suécia - dados epidemiológicos e curva ajustada.

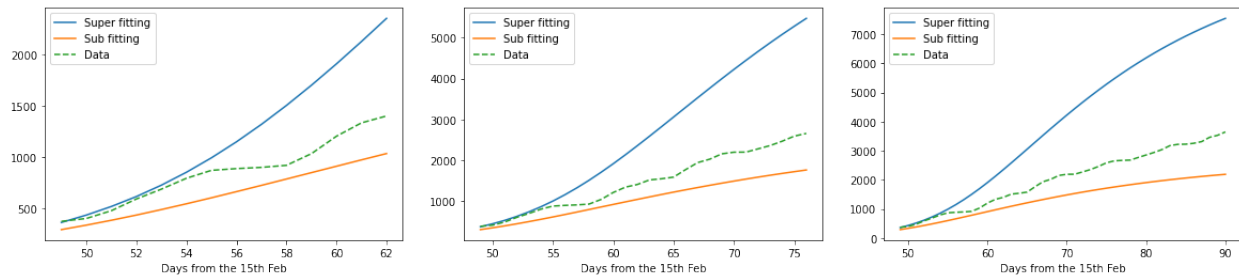


Figura 7: Suécia - Super / Sub-ajuste após o ajuste linear: 14 dias (esquerda); 28 dias (centro); 42 dias (direita).

Perspectiva de longo prazo

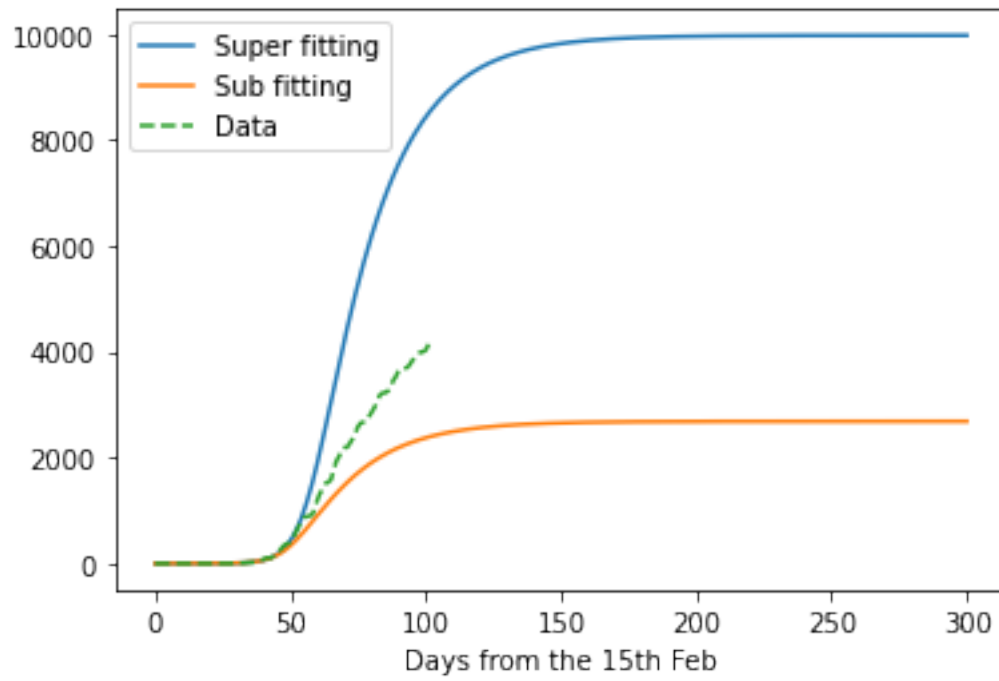


Figura 8: Suécia - A evolução "completa" das curvas de bracketing, juntamente com os dados disponíveis até o momento.

USA

- $\lambda_1 = 0,273813158$ e $K_1 = 0,0280839988$
- Coeficiente de determinação $R^2 = 0,9913530374232924$
- Período de ajuste linear: 28 – 42 (15 dias)

	\mathcal{R}_0	I_0	δ	μ
Super-ajuste	6,75	307363	1467115	5×10^{-4}
Sub-ajuste	2,92	567	3,7	$2,2 \times 10^{-4}$

Tabela 3: EUA - parâmetros de ajuste.

Ajuste e bracketing lineares

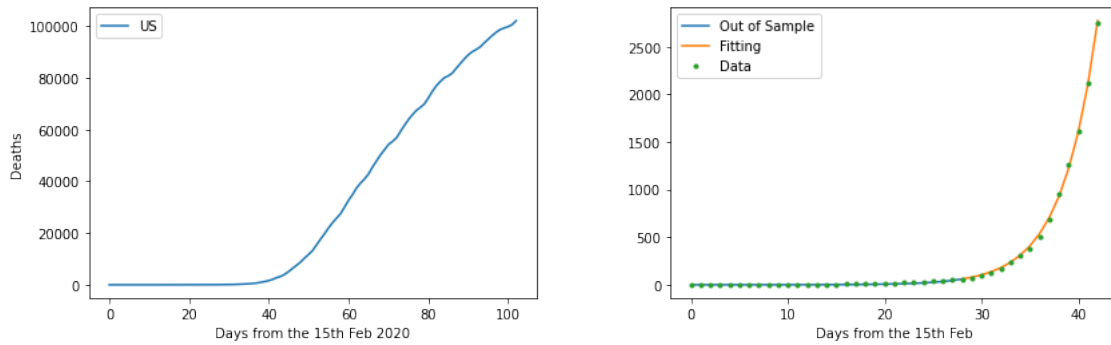


Figura 9: EUA - dados epidemiológicos e curva ajustada.

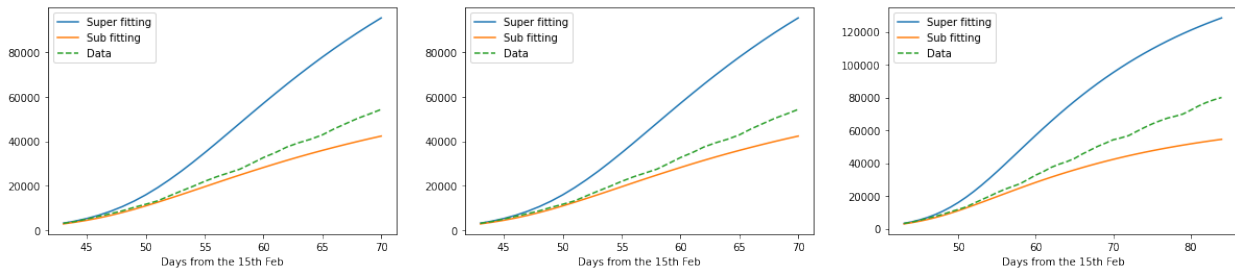


Figura 10: EUA - Super / Sub-ajuste após o ajuste linear: 14 dias (esquerda); 28 dias (centro); 42 dias (direita).

Perspectiva de longo prazo

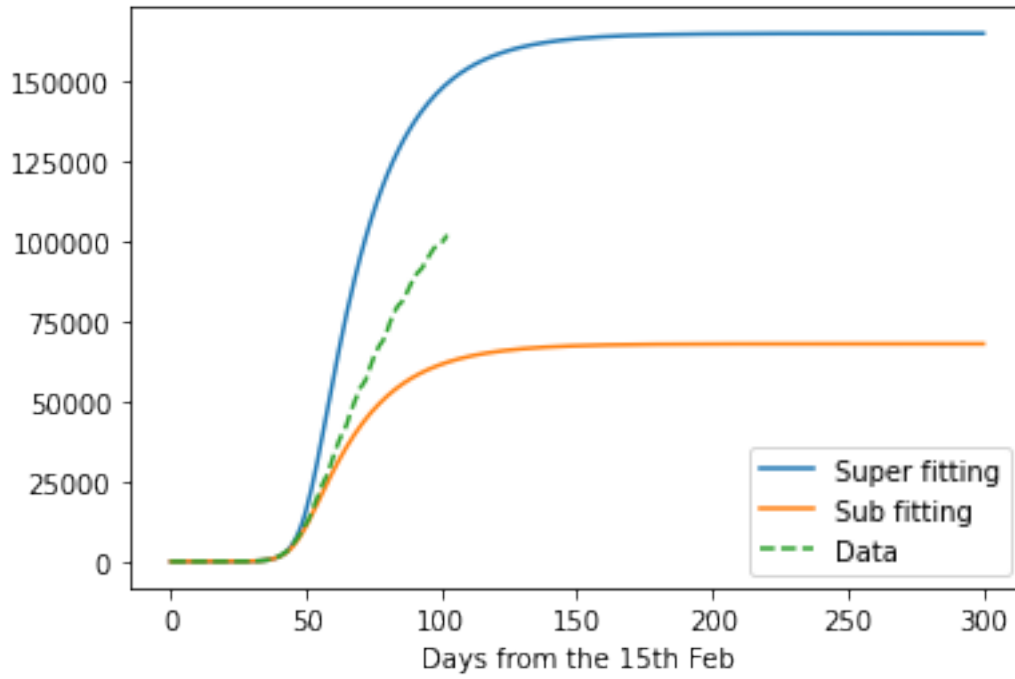


Figura 11: EUA - A evolução "completa" das curvas de bracketing, juntamente com os dados disponíveis até o momento.

Brasil

- $\lambda_1 = 0,194033766$ e $K_1 = 0,322620196$
- Coeficiente de determinação $R^2 = 0,9934307717963164$
- Período de ajuste linear: 39 – 47 (9 dias)

	\mathcal{R}_0	I_0	δ	μ
Super-ajuste	5,07275	625681	3411458	2×10^{-4}
Sub-ajuste	2,361	2698	21,54	$3,5 \times 10^{-5}$

Tabela 4: Brasil - parâmetros de ajuste.

Ajuste e bracketing lineares

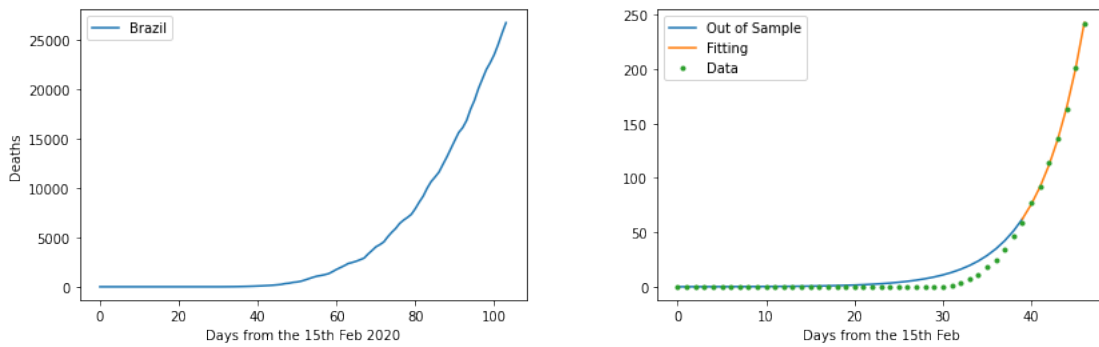


Figura 12: Brasil - dados epidemiológicos e curva ajustada.

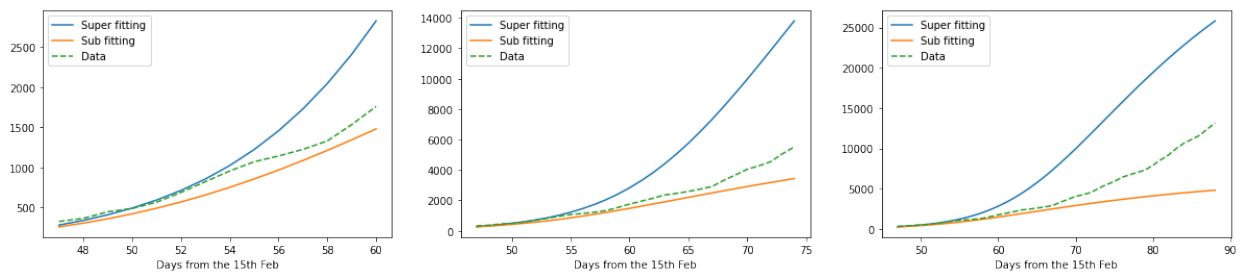


Figura 13: Brasil - Super / Sub-ajuste após o ajuste linear: 14 dias (esquerda); 28 dias (centro); 42 dias (direita).

Perspectiva de longo prazo

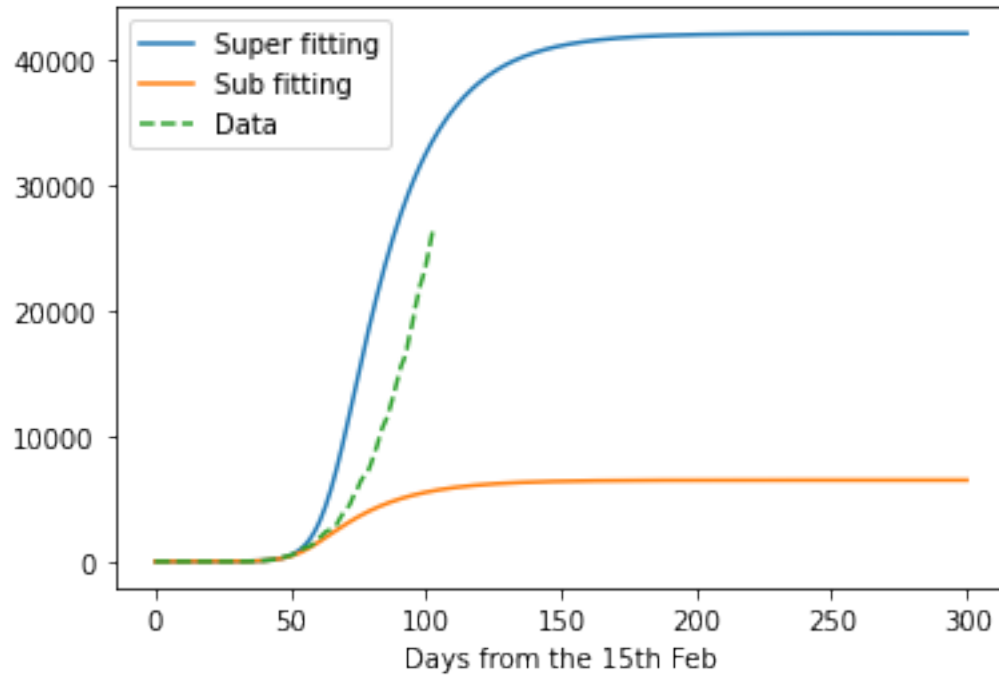


Figura 14: Brasil - A evolução geral das curvas de bracketing, juntamente com os dados disponíveis até o momento.

Discussão

Neste boletim (e no próximo que o complementa) procuramos entender quais informações sobre a doença podem ser extraídas da série temporal de óbitos em uma epidemia com dinâmica similar à covid-19.

No atual boletim examinamos a situação inicial da epidemia, onde os casos e óbitos crescem de maneira aproximadamente exponencial. Vimos que o R_0 pode ser determinado com razoável precisão, mas a incerteza sobre a letalidade é enorme, o que gera previsões de longo prazo sobre o total de óbitos extremamente incertas, na falta de informações auxiliares.

Apesar da falta de precisão na previsão da evolução dos óbitos, há limitações sobre o que pode ocorrer e o leque de cenários possíveis pode ser delimitado (*bracketing*) por dois cenários limites, o melhor e o pior cenário. A avaliação deste *bracketing* para dados de vários lugares foi feita com os dados de óbitos destas localidades. Os cenários analisados com dados reais devem ser vistos com muita cautela, pois a premissa de "população com mistura homogênea" presente em modelos SIR se enfraquece quando as regiões estudadas são muito grandes, como nos casos estudados. Portanto, os cenários descritos com base em dados reais tem antes a intenção de ilustrar o método do que fazer previsões para tais localidades.

Os dados de óbitos foram escolhidos no lugar do número de casos por consideramos que é mais frequente que os dados de óbitos tenham menos imprecisões, em particular quanto à subnotificação e/ou a limitação do número de testes. É claro que em uma localidade onde a série temporal do número de casos seja confiável, o uso desta informação pode reduzir bastante às incertezas das previsões.

Outra fonte clara de informações que pode ser usada para mitigar as incertezas é a forma como a curva de óbitos se diferencia do crescimento exponencial, o que ocorre quando a epidemia avança. Mas ocorre que no caso da covid-19, quando a epidemia avançou na maioria dos lugares os parâmetros da dinâmica de transmissão da doença não permaneceram constantes ao longo do tempo. Em particular, o parâmetro β , relacionado à infecciosidade, pode ter mudado radicalmente com a adoção de novas normas de comportamento e medidas de isolamento social. Essa mudança no parâmetro introduz novas incertezas. Usando dados da fase linear e não linear temos, portanto, uma quantidade maior de informações mas, ao mesmo tempo, um aumento na incerteza quanto à variação parâmetro β com o tempo. Esse problema será tratado em um próximo boletim onde mostraremos que a não-identificabilidade dos parâmetros permanece, assim como a incerteza quanto à previsão do total de óbitos. A complementação dos dados de óbitos com outros dados, como o número de indivíduos que já foram infectados e o número de indivíduos que são naturalmente mais resistentes a infecção, poderia praticamente eliminar as incertezas atuais. Infelizmente, essas informações complementares têm se mostrado difíceis de serem obtidas.

Referências

- [1] Esteban Ortiz-Ospina Max Roser, Hannah Ritchie and Joe Hasell. Coronavirus pandemic (covid-19). Our World in Data, 2020. <https://ourworldindata.org/coronavirus>.
- [2] M. A. Silva, H. Schechtman, and M. O. Souza. A death based model for covid-19, 2020. Poster presented at eSMB2020.
- [3] Junling Ma. Estimating epidemic exponential growth rate and basic reproduction number. Infectious Disease Modelling, 5, 01 2020.
- [4] Yang Liu, Li-Meng Yan, Lagen Wan, Tian-Xin Xiang, Aiping Le, Jia-Ming Liu, Malik Peiris, Leo L M Poon, and Wei Zhang. Viral dynamics in mild and severe cases of COVID-19. The Lancet Infectious Diseases, 20(6):656–657, June 2020.