

Mathematical Modeling

传染病模型 (第四次作业报告)

2024 年 5 月 5 日

马天开

tiankaima@mail.ustc.edu.cn

ID: 15 / PB2100030

摘要 / Introduction

本次作业中使用 Python 实现了:

- SIR 模型的数值解法
- SEIR 模型的数值解法
- 基本传染数 - 最终感染比例 $\mathcal{R}_0 - R_\infty$ 的计算
- Covid-19 早期数据拟合

我们将在本文中讨论如下内容:

- 传染病模型的发展历程
- SIR 模型
 - 模型的基本假设
 - 动力学方程
 - Explicit Euler 方法
 - Implicit Euler 方法
 - 对疫情发展的预测
 - 基本传染数 \mathcal{R}_0
 - R_∞ 与 \mathcal{R}_0 的关系
- SEIR 模型
 - 不同参数对疫情发展的影响
 - 潜伏期 $1/\sigma$
 - 康复率 γ
- 数据拟合
 - Covid-19 早期数据拟合
 - 潜伏期、康复率、死亡率的推断
 - 解释 Covid-19 大流行期间的困难, 可能的成因、解决方案

在本次作业报告中, 我们将从简单的 SIR 模型开始, 逐渐引入更加复杂的 SEIR 模型, 并讨论不同参数对疫情发展的影响. 我们将使用 Python 实现这些模型, 并通过数值方法求解这些模型的动力学方程.

前言 / Background

传染病模型

传染病研究开始于 Daniel Bernoulli 在 1760 年对天花疫苗的研究。传染病模型是数学建模中的一个重要领域，对于预测疫情的发展趋势、制定防控措施、评估防控措施的效果等方面有重要意义。

传染病模型从最开始的 SI 模型，到后来的 SIR 模型、SEIR 模型等，逐渐完善。传染病模型的建立需要考虑传染病的特性，如潜伏期、二次感染、死亡率等。在接下来的讨论中，我们只关注疫情早期的发展，在此基础上我们忽略二次感染的问题。因此不分析诸如 SIS SEIRS 等考虑 Recovered 重新变回 Susceptible 的模型。

根据研究方法的不同，传染病模型可以分为差分模型和连续模型，前者是后者的离散形式。差分模型通常用于疫情的传播过程的离散模拟，而连续模型通常用于疫情的传播过程的连续模拟。

动力学方程

传染病模型的动力学方程是一组常微分方程，描述了易感者、感染者和康复者的数量随时间的变化。传染病模型的动力学方程是一个非线性的方程组，通常难以求解。因此，我们需要使用数值方法来求解传染病模型的动力学方程。

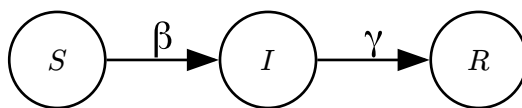
一般的动力学方程有如下的数值解法：

- Euler 方法: 包含显式 (Explicit) 和隐式 (Implicit) 的 Euler 方法
- Runge-Kutta 方法: 包含 2 阶、3 阶、4 阶的 Runge-Kutta 方法，常用的是 4 阶 Runge-Kutta 方法
- 中点 (Midpoint) 方法: 2 阶 Runge-Kutta 方法的一种特例

数学模型 / Mathematical Model

SIR 模型

传染病模型的一个经典模型是 SIR 模型。SIR 模型是用三个变量描述传染病的传播过程：S 表示易感者 (Susceptible)，I 表示感染者 (Infectious)，R 表示康复者 (Recovered)。SIR 模型的动力学方程是一组常微分方程，描述了易感者、感染者和康复者的数量随时间的变化。



这种简化模型主要适用于以下特征的传染病：

- 无潜伏期或者潜伏期很短
- 不会二次感染
- 不考虑超额死亡率

用以下常微分方程来表示：

$$\begin{cases} \frac{dS(t)}{dt} = \alpha \cdot S(t) - \beta \cdot S(t)I(t) - \mu \cdot S(t) \\ \frac{dI(t)}{dt} = \beta \cdot S(t)I(t) - \gamma \cdot I(t) - \mu \cdot I(t) \\ \frac{dR(t)}{dt} = \gamma \cdot I(t) - \mu \cdot R(t) \end{cases}$$

其中 α 代表种群的招募率(出生率), β 代表感染率, γ 代表康复率, μ 代表自然死亡率, 均为正参数. 注意这里定义中 S, I, R 均取比例, 即 $S + I + R = 1$; $\hat{\beta} = \beta/N$ 也已正交化, 与一般的定义有所不同.

Euler 方法

我们以 SIR 模型为基础, 先简单讨论如何对动力学方程进行离散化:

考虑初始条件 $S(t) = S_0, I(t) = R(t) = 0$, 我们可以使用 Euler 方法进行离散化:

$$\begin{aligned} S(n+1) - S(n) &= (\alpha S(n) - \beta S(n)I(n) - \mu S(n)) \cdot h \\ I(n+1) - I(n) &= (\beta S(n)I(n) - (\gamma + \mu)I(n)) \cdot h \\ R(n+1) - R(n) &= (\gamma I(n) - \mu R(n)) \cdot h \end{aligned}$$

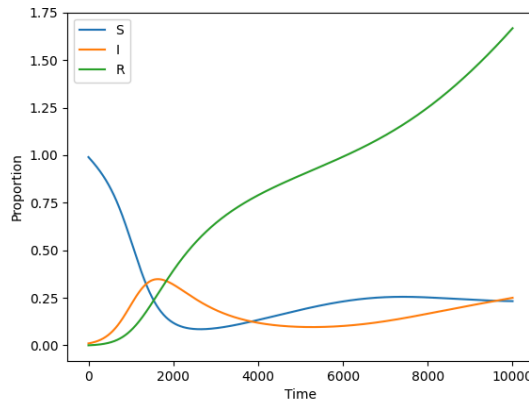
这实际上给出了一个通过 $S(n), I(n), R(n)$ 直接计算 $S(n+1), I(n+1), R(n+1)$ 的方法, 上述离散化的方法也被称为 Explicit Euler 方法.

作为参考, 我们也给出在这里使用 Implicit Euler 方法的离散化:

$$\begin{aligned} S(n+1) - S(n) &= (\alpha S(n+1) - \beta S(n+1)I(n+1) - \mu S(n+1)) \cdot h \\ I(n+1) - I(n) &= (\beta S(n+1)I(n+1) - (\gamma + \mu)I(n+1)) \cdot h \\ R(n+1) - R(n) &= (\gamma I(n+1) - \mu R(n+1)) \cdot h \end{aligned}$$

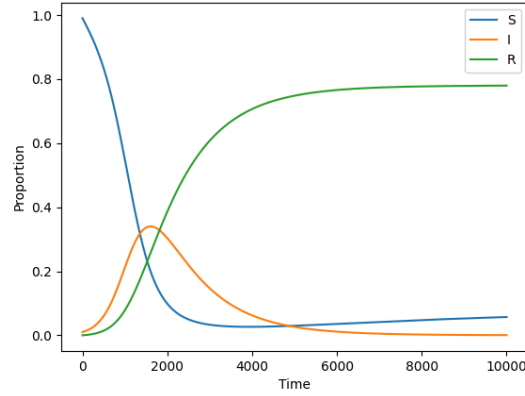
一般来说, 在这之后需要计算一个 $S(n+1), I(n+1), R(n+1)$ 的迭代方程, 通过迭代方法来求解. 在这个问题中, 使用 Explicit Euler 带来精度的提升是非常有限的(与直接降低 Explicit Euler 中迭代步长 h 相比). 但是, 在一些情况下, Implicit Euler 方法可能会更加稳定. (通常是更高维、线性问题)

实现上述 Explicit Euler, 我们得到:



$$\begin{aligned} \alpha &= 0.02, \quad \beta = 0.5, \quad \gamma = 0.1, \quad \mu = 0.01 \\ S_0 &= 0.99, \quad I_0 = 0.01, \quad R_0 = 0 \end{aligned}$$

调整出生率到合理范围:



$$\alpha = 0.002, \quad \beta = 0.5, \quad \gamma = 0.1, \quad \mu = 0.01$$

$$S_0 = 0.99, \quad I_0 = 0.01, \quad R_0 = 0$$

$$R_\infty(\mathcal{R}_0) = R_\infty(\beta, \gamma)$$

参考上面模拟结果, 我们继续简化 SIR 模型, 考虑在传染病爆发的初期, 忽略自然出生率、死亡率. 这样我们可以得到一个更简单的 SIR 模型:

$$\begin{cases} \frac{dS(t)}{dt} = -\beta \cdot S(t)I(t) \\ \frac{dI(t)}{dt} = \beta \cdot S(t)I(t) - \gamma \cdot I(t) \\ \frac{dR(t)}{dt} = \gamma \cdot I(t) \end{cases}$$

我们接下来只关心最终感染比率 R_∞ 与传染率 β 和康复率 γ 的关系. 我们可以通过求解上述方程组得到 S_∞ 与 β, γ 的关系.

考虑约化时间 $\tau = \gamma \cdot t$, 约化常量 $\mathcal{R}_0 \equiv \beta/\gamma$:

$$\begin{cases} \frac{dS}{d\tau} = -\mathcal{R}_0 SI \\ \frac{dI}{d\tau} = \mathcal{R}_0 SI - I \\ \frac{dR}{d\tau} = I \end{cases}$$

考察这样简化后的不动点 (S_*, I_*, R_*) , 既然是不动点, 显然有:

$$\frac{dS}{d\tau} = \frac{dI}{d\tau} = \frac{dR}{d\tau} = 0$$

其中

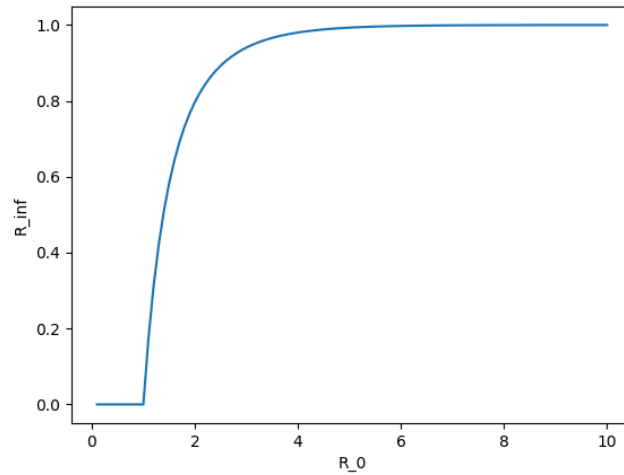
$$\frac{dR}{d\tau} = I = 0 \Rightarrow I_* = 0$$

显然此时 $(S_*, I_*, R_*) = (1 - R_\infty, 0, 1 - R_\infty)$

可以得到:

$$\begin{aligned}\frac{dS}{dR} &= \frac{\frac{dS}{d\tau}}{\frac{dR}{d\tau}} = -\mathcal{R}_0 S \\ \Rightarrow \int_{S_0}^{S_\infty} \frac{dS}{S} &= -\mathcal{R}_0 \int_{R_0}^{R_\infty} dR \\ \Rightarrow 1 - R_\infty - e^{-\mathcal{R}_0 R_\infty} &= 0\end{aligned}$$

这里我们得到了一个关于 $R_\infty(\mathcal{R}_0)$ 的超越方程, 可以通过程序绘制出 $\mathcal{R}_0 - R_\infty$:



我们观察到相当有趣的结果, 当 $\mathcal{R}_0 < 1$ 时, $R_\infty = 0$, 即传染病不会爆发; 当 $\mathcal{R}_0 > 1$ 时, R_∞ 随 \mathcal{R}_0 的增大而增大, 传染病会爆发.

我们接下来通过另一种方法引入 $\mathcal{R}_0 = \beta/\gamma$ 的定义, 读者很快就能理解上述结论的合理性.

基本传染数 \mathcal{R}_0

抛开上面的讨论, 我们假设一个零号病人, 记他在 t 时刻仍未康复的概率为 $l(t)$, 有

$$\frac{dl}{dt} = -\gamma l$$

考虑初值条件 $l(0) = 1$, 这是个基本的指数衰减问题, 解得

$$l(t) = e^{-\gamma t}$$

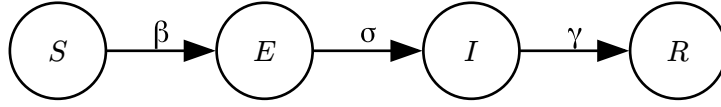
这个零号病人在生病期间内最多能感染多少人呢? 保持上面 β 的定义不变, 我们有:

$$\mathcal{R}_0 = \int_0^\infty \beta \cdot l(t) dt = \beta \cdot \int_0^\infty e^{-\gamma t} dt = \frac{\beta}{\gamma}$$

按照这个定义, \mathcal{R}_0 可以被理解为「平均每个感染者能感染的人数」, 也被称为**基本传染数**. 容易想到, 当 $\mathcal{R}_0 < 1$ 时, 传染病不会爆发; 当 $\mathcal{R}_0 > 1$ 时, 传染病会爆发.

SEIR 模型

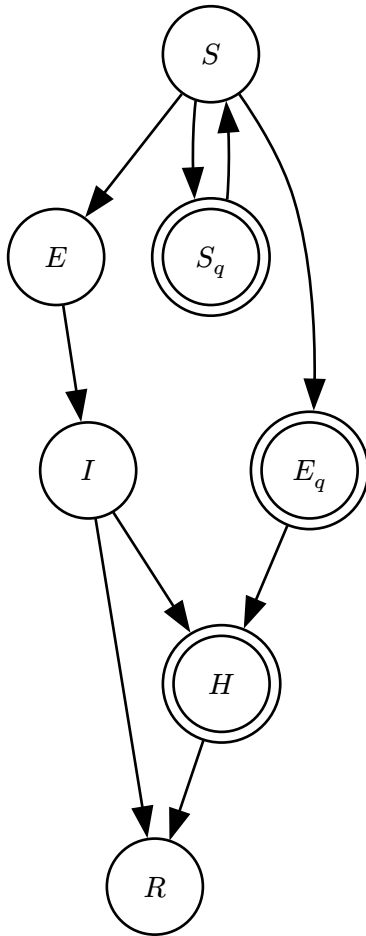
在 SIR 模型的基础上, 我们引入了潜伏者 (Exposed) 的概念, 得到了 SEIR 模型:



类似的, 我们可以得到 SEIR 模型的动力学方程: (为方便讨论, 我们省略时间参数, 但是读者应该注意到这里的 S, E, I, R 都是关于时间的函数)

$$\begin{cases} \frac{dS}{dt} = \alpha S - \beta SI - \mu S \\ \frac{dE}{dt} = \beta SI - \sigma E - \mu E \\ \frac{dI}{dt} = \sigma E - \gamma I - \mu I \\ \frac{dR}{dt} = \gamma I - \mu R \end{cases}$$

在 Covid-19 早期爆发的例子中, 上述模型实际上可以更加细致的表述为下面的形式:



$$\begin{aligned} \frac{dS}{dt} &= -[\rho c \beta + \rho c q(1 - \beta)]S(I + \theta E) + \lambda S_q \\ \frac{dE}{dt} &= [\rho c \beta(1 - q)]S(I + \theta E) - \sigma E \\ \frac{dI}{dt} &= \sigma E - (\delta_l + \alpha + \gamma_l)I \\ \frac{dR}{dt} &= \gamma_l I + \gamma_H H \\ \frac{dS_q}{dt} &= \rho c q(1 - \beta)S(I + \theta E) - \lambda S_q \\ \frac{dE_q}{dt} &= \rho c \beta q S(I + \theta E) - \delta_q E_q \\ \frac{dH}{dt} &= \delta_l I - (\alpha + \gamma_H)H + \delta_q E_q \end{aligned}$$

参考文献:

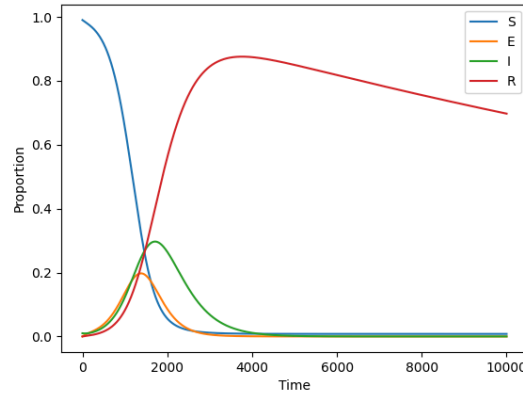
Cao S, Feng P, Shi P. Zhejiang Da Xue
Xue Bao Yi Xue Ban.
2020;49(2):178-184. doi:10.3785/
j.issn.1008-9292.2020.02.05

细致化的模型中, 引入了隔离者 S_q , 隔离的潜伏者 E_q , 以及医院患者 H . 这个模型更加细致的考虑了隔离、医院治疗等因素, 但我们在这里不详细讨论.

SEIR 模型的数值解法 - Implicit Euler

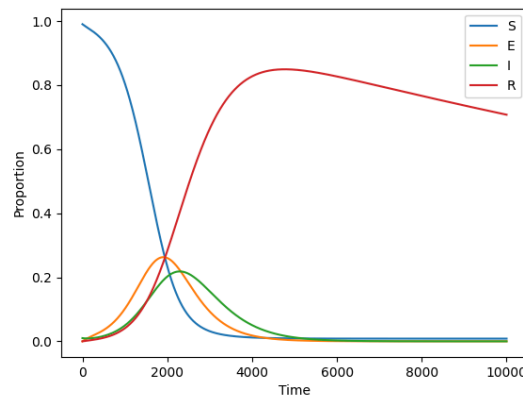
我们可以使用 Implicit Euler 方法对 SEIR 模型进行数值解法. 与 SIR 模型类似, 我们可以得到:

$$\begin{cases} \frac{dS}{dt} = \alpha S - \beta SI - \mu S \\ \frac{dE}{dt} = \beta SI - \sigma E - \mu E \\ \frac{dI}{dt} = \sigma E - \gamma I - \mu I \\ \frac{dR}{dt} = \gamma I - \mu R \end{cases} \Rightarrow \begin{cases} S(n+1) - S(n) = (\alpha S(n) - \beta S(n)I(n) - \mu S(n)) \cdot h \\ E(n+1) - E(n) = (\beta S(n)I(n) - \sigma E(n) - \mu E(n)) \cdot h \\ I(n+1) - I(n) = (\sigma E(n) - \gamma I(n) - \mu I(n)) \cdot h \\ R(n+1) - R(n) = (\gamma I(n) - \mu R(n)) \cdot h \end{cases}$$



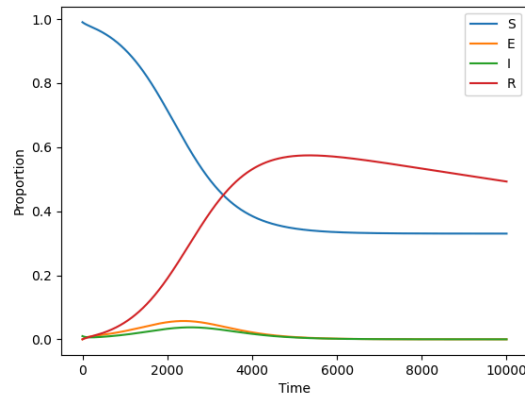
$$\alpha = 0.002, \quad \beta = 0.5, \quad \sigma = 0.2, \quad \gamma = 0.1, \quad \mu = 0.002$$
$$S_0 = 0.99, \quad E_0 = 0, \quad I_0 = 0.01, \quad R_0 = 0$$

延长潜伏期 $1/\sigma$, $5 \rightarrow 10$ 天, 我们可以得到:



$$\alpha = 0.002, \quad \beta = 0.5, \quad \underline{\sigma = 0.1}, \quad \gamma = 0.1, \quad \mu = 0.002$$
$$S_0 = 0.99, \quad E_0 = 0, \quad I_0 = 0.01, \quad R_0 = 0$$

增大 γ (康复率) 到 0.2, 我们可以得到:

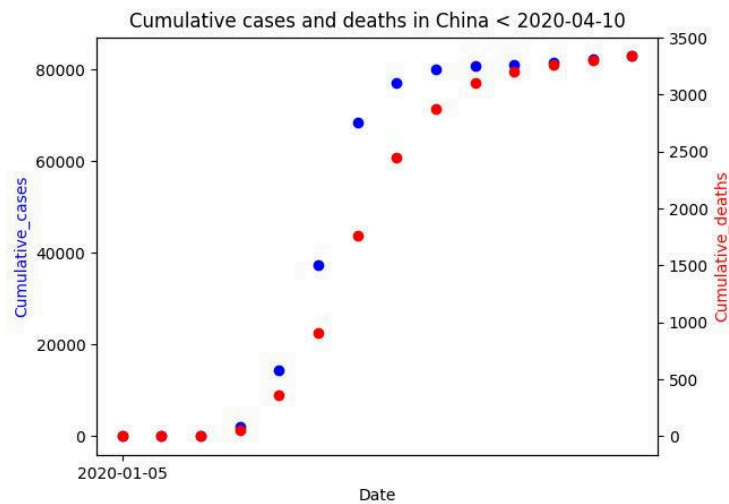


$$\alpha = 0.002, \quad \beta = 0.5, \quad \sigma = 0.2, \quad \underline{\gamma = 0.3}, \quad \mu = 0.002$$
$$S_0 = 0.99, \quad E_0 = 0, \quad I_0 = 0.01, \quad R_0 = 0$$

结果与对比 / Results and Comparison

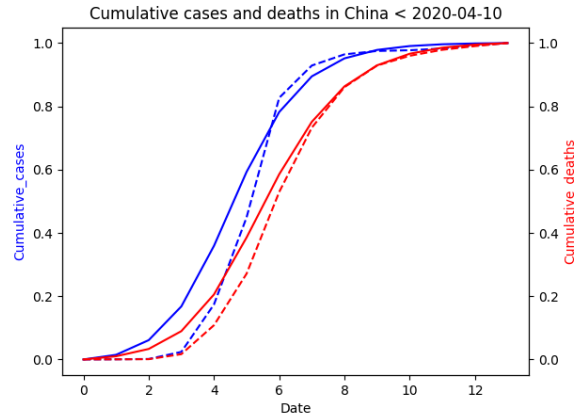
数据拟合

根据 WHO 提供的数据, 我们尝试拟合一组 $(\beta, \sigma, \gamma, \dots)$ 的参数, 使得模型与实际数据相符; WHO 提供的数据如下: (中国, 早于 2020 年 4 月 1 日, 对应疫情早期)



考虑到统计口径只能记录确诊人数 K 、死亡人数 D , 我们从原 (S, E, I, R) 模型中提取这两者:

$$\begin{cases} \frac{dK}{dt} = \sigma E \\ \frac{dD}{dt} = \delta_I I \end{cases} \Rightarrow \begin{cases} K = \int_0^t \sigma E dt \\ D = \int_0^t \delta_I I dt \end{cases}$$



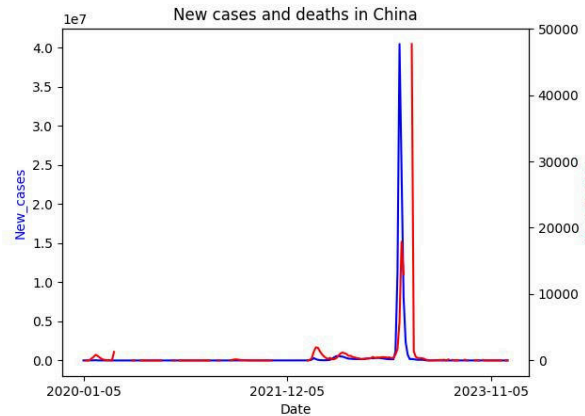
$$\alpha = 0.002, \quad \beta = 0.5, \quad \gamma = 0.09, \quad \mu = 0.002, \quad \sigma = 0.14, \quad \sigma_I = 0.03, \quad S_0 = 0.99$$

(虚线: 真实数据, 实线: 模型拟合; 红线: 累计确诊人数, 蓝线: 累计死亡人数)

按此拟合数据推断, 早期 Covid-19 潜伏期 $1/\sigma = 7$ 天, 康复率 $\gamma = 0.09$, 死亡率 $\sigma_I = 0.03$, $\mathcal{R}_0 = \beta/\gamma = 5.56$, 与实际数据相符.

结论 / Conclusion

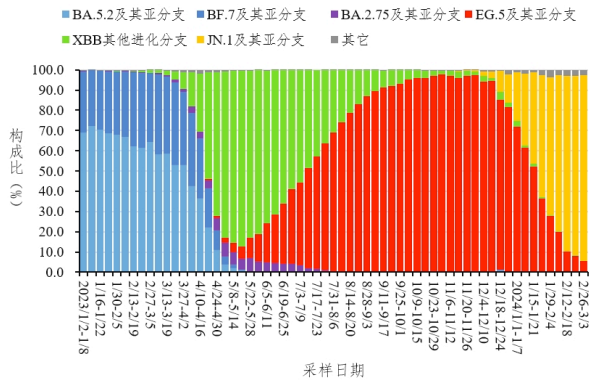
拟合整个区间的困难



SEIR 模型不能成功拟合整个 CN 疫情的数据, 模型大面积简化了隔离、医院治疗等情况, 忽略这些因素随时间变化、医院收治能力上限、政策因素等; 在大流行期间更重要的问题, 例如变种、免疫逃逸、疫苗接种等, 也没有考虑.

可能的解决方案

- 引入更多的参数, 考虑隔离、医院治疗等因素
- 考虑疫苗接种、免疫逃逸等因素
- 考虑变种对疫情发展的影响
 - 将每个变种视为一个单独的传染病模型, 考虑变种间的传播
 - 重新评估每个变种传播前易感人群的数量 S_0, R_0
- 考虑政策因素对疫情发展的影响
 - 将部分因子作为时间变量, 考虑政策因素对 β, σ, γ 等的影响

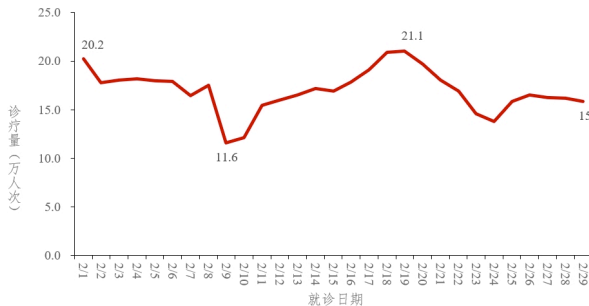


中国疾病预防控制中心 (CDC) 对 Covid-19 变种随时间的变化的统计

在此基础上, 为增大模型拟合的精度, 可以考虑诸如污水监测、病毒基因测序等更加细致的数据(获得更多的参数), 以及更加复杂的模型.

现实意义

传染病模型为疫情早期的预测提供了重要的参考, 这直接影响了防控措施、时长, 也科学有效地向大众说明了疫情的发展规律. 但是, 传染病模型的局限性也是显而易见的, 例如在大流行期间, 传染病模型的复杂性、不确定性、政策因素等都会对模型的精度产生影响.



中国疾病预防控制中心 (CDC) 对发热门诊接诊量随时间的变化的统计

然而限于统计方法、数据收集方面等障碍, 传染病模型的精度仍然有待提高. 传染病模型的发展也是一个不断完善的过程, 未来的研究将会更加注重模型的复杂性、数据的准确性等.

备注 / Remarks

- 安装:
`pip install numpy matplotlib pandas scipy`