# 面向COVID-19的传播动力学建模

## 1. 传播动力学模型

### 1.1 基本传播动力学模型SIR

经典SIR模型[1]将群体状态分成三类：易感者（S）、感染者（I）、康复者（R），其中易感者是指没有被感染但缺乏免疫能力并且容易被感染者传染的群体；感染者是指已经被感染并且能够将病毒传染给易感者的群体；痊愈者是指已经隔离或者对病毒具有免疫能力的群体。这三类群体的的状态按照一定的转换率互相转换，转移图如图1所示，构建对应的SIR模型的传播动力学方程式（1）。



图1 经典SIR模型的状态转移图

(1)

式中S(t)、I(t)、R(t)分别是t时刻的易感者、感染者与康复者的人数；N为人口总数；β表示接触传染率；γ表示感染者的治愈率。

### 1.2 引入人口流动的SEIR模型

针对COVID-19这种疾病，传统的SIR模型并不能够反映实际的传播特点，结合COVID-19具有一定的潜伏期的问题，并引入人口的流入流出，来体现疾病的扩散风险。在此基础上我们使用改进的SEIR模型，该模型引入了潜伏期的个体类型，并将人口的流入流出情况纳入考虑，改进后的SEIR模型如图2所示

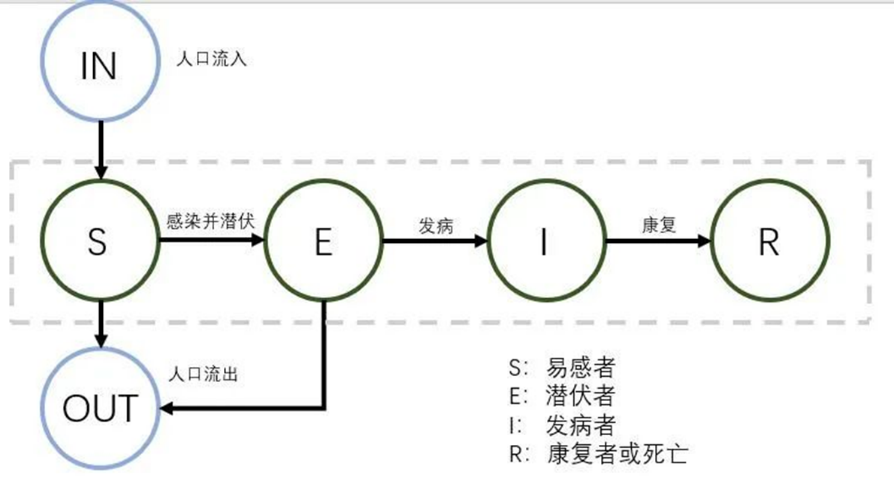


图2 引入人口流动的SEIR模型

由于人口的流动与传染病的各个状态的相关性较弱，因此我们将IN与 OUT设定为外部量，考虑到IN与OUT都是在离散时间上的变量，我们可以得到引入人口流动的SEIR模型的传播动力学的差分方程

(2)

其中S(t),E(t),I(t),R(t),IN(t),OUT(t),N(t)分别代表t时刻易感者人数、潜伏者人数、发病者人数、康复者人数、流入人口数量、流出人口数量,总人口数量，这些变量基本都能由现实的数据直接或者间接的获得，是模型推理的数据基础。

### 1.3 SEIR模型的简化

在以上模型的基础上为了计算能进行，我们需要根据现实情况对模型进行一定的简化：1. 由于武汉最高峰感染人数60000，相较于一千万的人口比例较小，因此将S(t)视作与N(t)一致的值；2. 潜伏期的人数相较于易感人群的数量也极小，因此流出人口对于潜伏期人数的影响也可忽略不计。因此简化后的模型如下所示：

(3)

### 1.4 参数辨识模型

上述模型中α、β、γ分别代表潜伏者转变成感染者的速率，感染者疾病传染的速率，感染者康复或被的速率，这些数据里部分能通过现实的病理学研究直接获取，部分只能通过参数辨识问题求解。我们可以得到所有的参数如下表

|  |  |
| --- | --- |
| 参数名称 | 解释 |
| S(t) | t时刻易感者人数 |
| E(t) | t时刻潜伏者人数 |
| I(t) | t时刻感染者人数 |
| R(t) | t时刻痊愈者或被收入方舱的人数 |
| IN(t) | t时刻流入人口数量 |
| OUT(t) | t时刻流出人口数量 |
| N(t) | t时刻总人口数量 |
| α | 潜伏者转变成感染者的速率，其倒数为潜伏期长度 |
| β | 感染者传染的速率 |
| γ | 感染者康复的速率 |

表1 SEIR模型中参数意义

由于潜伏期长度与感染者康复的速率并不由人为干预改变，因此，我们可以以现实的回顾性病理研究为准，获取该数据，而根据[2]中的数据，可以得知，平均潜伏时间为5.2d，平均康复时间为10d。而β与实际的防控措施相关，因此即使我们有病理学的研究数据，依然只能通过参数辨识问题来求解。因此我们以\*代表实际数据，可以得到以下参数辨识问题。

(3)

而上述优化问题可以借助L-M方法求解最佳参数估计。

### 1.5 模型精度估计

将模型的计算结果与真实的疫情数据进行比较，可以验证模型的模拟精度。在这里定义模型计算结果与真实数据之间的平均偏差为模型的模拟误差e，表达式如下所示

式中T代表真实数据包含的天数，模型的模拟精度e越小，表示模拟结果与真实数据越接近，模型的拟合精度与预测精度越高，模拟效果越好。

## 2. 武汉疫情的分析

### 2.1 基于政策的阶段分析

根据以上的分析，我们可以知道决定武汉疫情中参数的确定的关键因素为政府的防控措施，因此我们根据政府的防疫措施将疫情分成三个阶段：

第一阶段--无措施期（2019/12/27-2020/1/22）：该阶段政府没有采取任何有效的防控措施。

第二阶段--封城初期（2020/1/23/-2020/2/15）:该阶段由于钟南山院士宣布存在“人传人”现象，武汉开始封城，市民活动强度骤减，政府大幅度提高病例检测力度，但由于患者数量激增，武汉市医疗资源被挤兑，大量确诊病例无法住院，只能回家自我隔离或不断赴医院排队，造成大量家庭和医院感染。

第三阶段--全面严控期（2020/2/15-2020/3/19）:政府采取更严格管控措施限制人员活动和密切接触；紧急增加床位并设立方舱医院，力推“大排查”和“应收尽收”，集中收治轻症患者，切断家庭和医院感染途径。

### 2.2 根据实际数据的分段分析

而由于模型中只考虑一个参数beta，因此函数理论上是分段的指数函数，想要找到最佳的拟合，需要根据数据具体情况进行分段。根据实际代码调试和数据函数形状，我们可以根据三个阶段的时间划分，分别计算最佳的参数估计，拟合程度最佳的函数分段应该为：

（设时间从2020/1/1开始，初始感染人数和潜伏期人数都设为1）

第一阶段为（2020/1/30之前）

β=1.0623，e=51，函数拟合图像为：



图1 第一阶段的实际数据和模型拟合数据

第二阶段为（2020/1/31-2020/2/8）

β=0.0710，e=129.88，函数拟合图像为：



图2 第二阶段的实际数据和模型拟合数据

第三阶段为（2020/2/9之后）。

β=-0.097，e=133.85，函数拟合图像为：



图3 第三阶段的实际数据和模型拟合数据

从数据出发的分段不同于从现实政策出发的分段的原因是多样的，一个是新冠具有潜伏期，疾病的感染数据有一定的时滞效应；一个是现实防疫政策的具体实施不同于官方声明的时期，武汉初期的防疫情况比较混乱；最后一个是武汉的检测能力有限，数据并不总是与真实情况一致。这些原因共同导致了我们只能从实际的数据出发去度量数据层面的武汉疫情情况。

通过整合三个阶段数据，我们可以得到完整的武汉疫情数据模拟：



图4 武汉疫情的实际数据和模型拟合数据

其中数据来源于武汉公共数据开放平台。

[1]. Dietz K. The first epidemic model: a historical note on PD En'ko[J]. Australian Journal of Statistics, 1988, 30(1): 56-65.

[2].中华预防医学会新型冠状病毒肺炎防控专家组.新型冠状病毒肺炎流行病学特征的最新认识[J].中国病毒病杂志,2020,10(02):86-92.DOI:10.16505/j.2095-0136.2020.0015.