

# 模拟新冠病毒群体免疫的数学建模及数据分析

## 一、摘要

2020 年，新型冠状病毒在全球范围内引发疫情，近来，多个国家相继开始施行大规模接种疫苗，以建立免疫屏障，有效防控疫情，实现群体免疫。

针对问题一，我们用一个类沙盒模型——元胞自动机来模拟新冠病毒的传染过程。沙盒是一个在计算机中的完全隔离于外界的环境，在沙盒中的程序不能运行于沙盒之外，外界的任何变化与沙盒内无关[1]。为了尽可能简化模型以及逼近真实情况，我们把一群易感染人群和新冠携带者们随机混合在这个黑匣子中，建立其独特的传染机制及随机化的参数扰动，完成了随机参数元胞自动机的构建。使用机器学习得到线性相关的参数，确定了模型的走向。我们在其中设置了一系列的事件，通过这些事件改变相关状态变化的概率，来模拟社交隔离等情况。以此将疫苗的有效性和病毒的变异作为可调节的参数。创造性地使用 GPU 多线程增加训练速度，从而大大缩短训练时间。在此基础上，调试模型参数这一原本需要等待 108s 的过程被缩短到了五秒内，提高了模型更换参数的效率。

此时我们已经初步建立了传染病毒群体的数学模型，只需要在其中增加策略即可完成对免疫屏障的模拟。我们用一个规则矩（见图 6-1-2）存储所有的概率。

对于问题二，我们选取杭州为例。首先，我们使用了第一问所用的模型去模拟杭州在疫情期间的病毒传播的情况并建立出模型曲线。再使用差分法，得出图 6-2-1，区县中的前后两个峰值分别代表病毒传染强度较弱达到群体免疫的时间和疫苗强度较强时达到群体免疫的时间。取最概然达到群免的步长在第 58 步，用该步长模拟达到群体免疫的天数，由此计算成本。

对于问题三，我们搜集了美国 2020 年 1 月 1 日年到 2021 年 5 月 1 日各大州的疫情数据，并多次利用回归分析，使用 sigmoid 函数建立横纵坐标的关系，以方程

$$N_{Case} = \frac{k_0}{1 + e^{-(k_1 x + k_2)}} + k_3$$
高准确度地拟合了美国疫情期间累计确诊人数和累计死亡人数，与实际情况对照，得出施行防疫措施对控制美国疫情的影响。证明了疫苗接种对美国疫情控制的积极影响，由此并预测了美国疫情未来趋势的走向。

**关键词：**元胞自动机 新冠疫情预测 群体免疫 最优策略 基本再生数

## 二、问题重述

### 2.1 问题背景

为了巩固国内经济增长基础，保障国际国内双循环经济格局的发展，中国政府在全国范围内实施新冠病毒疫苗接种工作，以此建立最大规模的国民群体新冠病毒免疫屏障。

### 2.2 问题提出

基于以上背景我们需要解决以下问题：

- (1) 建立传染病毒群体免疫屏障的数学模型，借此说明疫苗接种率对构建免疫屏障发挥的作用，并基于此考虑疫苗的有效性问题和病毒的变异问题对免疫屏障可能带来的影响。讨论物理隔离和免疫屏障在对于整个社会的代价和收益问题上的异同。
- (2) 选取一个中国的大城市，以此为例，思考需要采取怎样的措施，需要多少成本和时间可完成群体免疫的屏障建立。
- (3) 作为世界上新冠疫情最严重的国家，美国也是大规模、快速度接种新冠疫苗的国家之一。近来美国的疫情趋势有所好转，试分析接种疫苗对美国疫情变化的影响并对美国疫情的走向做出预测。

## 三、问题分析

针对问题一：考虑到病毒在人群中的传染性，疫苗接种后人群的免疫性，我们可以根据个人当前的感染状况进行分类，其未来的健康状况的改变则由个人的接触者、病毒的传染、疫苗的接种、个人的是否是康复患者等因素有关。基于状态上时间上空间上都离散的变化，我们采用元胞自动机模型对其进行模拟。以模型演化中每一步的状态改变模拟疫情中的每一天各类人群的感染情况变化。针对疫苗的有效性和病毒的变异性问题，可通过调节治愈率、确诊率、死亡率、接种率等使模型更接近实际情况。

针对问题二：当我们建立好第一问所示的模型后，我们在此基础上加入了计算成本和回归的逻辑。事实上，要计算的成是与时间程线性相关的，我们定义了一个  $1 \times 5$  的矩阵去存储所需要的所有参数，如社交隔离带来的平均每人的  $x_0$  元损失，每只疫苗耗费  $x_1$  元，每位病患所需医疗补助  $x_2$  元，每死亡一个人损失了  $x_3$  元，核算检测  $x_4$  元。随后我们对每个事件进行计数：如果有人患病了，则对应的计数器增加上患病人数；如果有人死亡，则死亡损失的计数器加一。最后是向量点乘，拿到所有的值进行加权求和，得到最概然成本。

针对问题三：题目要求分析美国大规模快速接种疫苗对疫情的影响并对疫情未来走向做出预测。这就要求我们收集美国各州在接种前后疫情数据，发现接种前后疫情发展变化，利用相似曲线进行拟合，发现规律，计算确诊人数、死亡人数、健康人数的权重，利用机器学习中的线性回归模型对其未来趋势进行预测。

## 四、模型假设

1. 假设接种疫苗前后所有人对病毒的免疫程度都相同，且疫苗接种后不再被

COVID-19 感染。

2. 假设疫苗的作用是即时的。
3. 假设所有确诊者都已被隔离。
4. 假设不同地区人口密度均同，且每个人的活动半径都相同，每个人只会在本身所在范围内被感染。

## 五、符号说明

符号	说明	单位
$R_0$	最小再生数	
$P_{12}$	携带者的每步确诊概率	
$N_{case}$	美国新冠病毒感染累计确诊人数	百万人
$k_0$	回归模型的参数	
$k_1$	回归模型的参数	
$k_2$	回归模型的参数	
$k_3$	回归模型的参数	
$P_{01}$	易感者感染基础概率	
$x_0$	社交隔离每人每天带来的 GDP 损失	元/天
$x_1$	每一支疫苗的花费	元
$x_2$	每位病患每天的医疗补助	元/天
$x_3$	死亡一人的损失	元
$x_4$	核算检测费用	元
$W_{cost}$	总成本	元

## 六、模型的建立与求解

### 6.1 问题一的模型建立与求解

#### 6.1.1 背景知识

元胞自动机是由元胞、元胞空间、元胞邻居和元胞规则组成。元胞自动机并非由严格的函数或物理方程确定，而是有离散时间、离散空间、有限状态上按照一定规则进行演化的动力学系统。元胞又称单元、细胞或基元；元胞空间是元胞所分布在欧几里得空间上的网格点的集合，一般情况下按照三角形、四边形和六边形三种网格排列；元胞邻居指某一元胞更新时进行搜索的空间域网格排列；元胞规则指根据元胞当前状态和邻居状态确定下一时刻该元胞状态的动力学函数。

#### 6.1.2 模型建立

分析新型冠状病毒引发疫情期间人群组成的状态，我们发现人群可以分为：健康者、携带者、确诊者（已被隔离）、康复者（已形成抗体）、死亡者。利用病毒在人群内部的传染性和疫苗接种人群对病毒的免疫性，我们借助元胞自动机模型，建立类元胞自动机模型模拟个人在疫情期间的状态。

状态编号	状态描述
0	易感人群
1	COVID-19 携带者
2	COVID-19 确诊者
3	康复者（已形成抗体）和疫苗注射者
4	死亡者

表 5-1-1

起始时期，我们在 1024\*1024 的矩阵地图上随机放置一定量的 COVID-19 携带者、确诊者以及已产生抗体的康复者和疫苗注射者。总计模拟 120 步，以此模拟 120 天内各类人群的状态变化。在每步模拟中，元胞的状态变化如下表：

后状态 前状态	0	1	2	3	4
0	由其他格计算得出	$[1-(1-P_{01})^N]$	0	$P_{03}$	0
1	0	$P_{11}$	$P_{12}$	0	$P_{14}$
2	0	0	由其他格计算得出	$P_{23}$	$P_{24}$
3	0	0	0	1	0
4	0	0	0	0	1

表 6-1-2

注： $P_{01}$  (易感人群转化为携带者的概率)

$P_{12}$  (携带者的每步确诊概率)

$P_{03}$  (疫苗每步的接种概率)

$P_{23}$  (确诊者的每步治愈率)

$P_{14}$  (携带者的每步死亡概率)

$P_{24}$  (确诊者的每步病死率)

每一个元胞的空间域是以该元胞为中心的 5\*5 矩阵，即该元胞的状态受该 5\*5 矩阵中的其余每一个元胞邻居的影响，示意图如下：

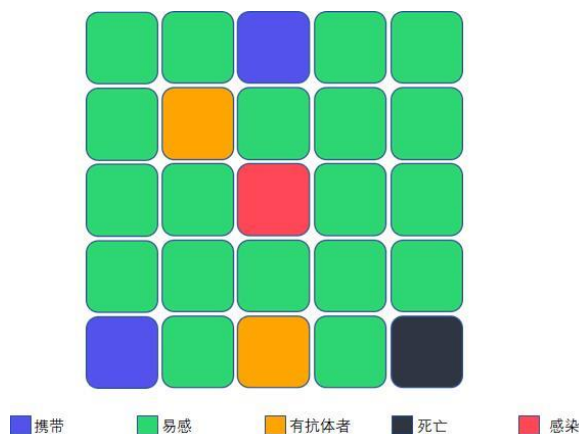


图 6-1-1

基于实际条件，针对疫苗的有效性和病毒的变异性，我们将其变为可调节的参数，并规定：

- 1、当确诊人数多于某个值时，将会施行社交防护措施，确诊率  $P_{12}$  下降，疫苗接种率  $P_{03}$  上升，这里设该值为 700。
- 2、当确诊人数大于某个值时，由于医疗供应所限，疫苗注射者的概率  $P_{03}$ 、携带者确诊率  $P_{12}$ 、确诊者治愈率  $P_{23}$  被迫下降，确诊者病死率  $P_{24}$  和携带者死亡率  $P_{14}$  随之升高，这里设该值为 1500。

### 6.1.3 模型的求解

$R_0$  指一个感染到某种传染病的人，会把疾病传染给其他多少个人的平均数。新型冠状病毒病毒的  $R_0$  约为 2.56 [2]

在这个模型里，我们所设置的概率参数很大程度上取决于  $R_0$ ，等待我们去解决的是如何建立概率和  $R_0$  之间的关系。

首先，我们求解了平均步长。然后，我们为了方便问题的解决，定义了整数  $h$ 。解决  $P_{01}$  和  $R_0$  的方案用了如下的两个函数。其中  $h$  是一个常数，经过两次的近似计算得出  $q$  和  $R_0$  的最后公式如下：

$$n_{mean} = \sum_{k=1}^{\infty} k(1 - P_{12})^{k-1} \times P_{12} = \frac{1}{P_{12}} \quad (1)$$

$$\left\lceil \frac{1}{P_{12}} \right\rceil = h \approx 14 \quad (2)$$

$$P_{01} = 1 - \sqrt[h]{1 - \frac{R_0}{24}} \approx 0.008 \quad (3)$$

$$\sum_{k=0}^{24} k \binom{24}{k} (1-q)^k q^{24-k} = R_0 \quad (4)$$

根据以上规定和计算机随机设置的各类人群在矩阵地图上的分布，以不同颜色代表不同人群，得到部分如下模拟示意图：

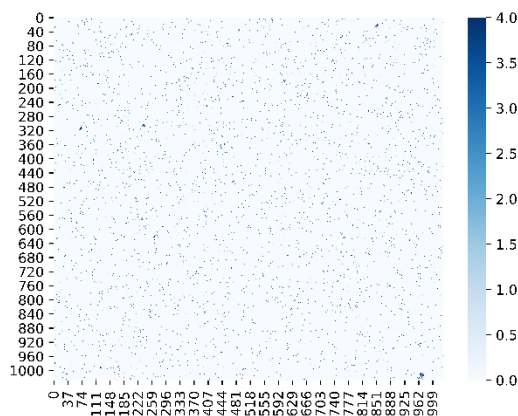


图 6-1-2 矩阵地图模拟示意图

分析模拟结果，得到模拟结果的数据统计图：

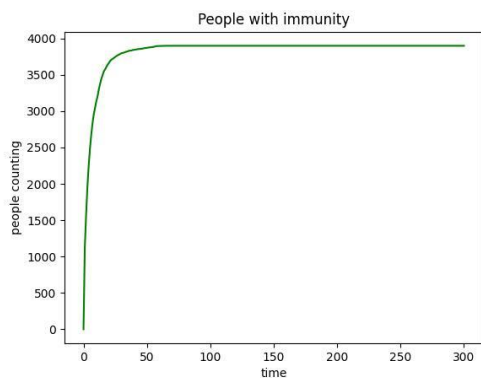


图 6-1-3 抗感染者人数

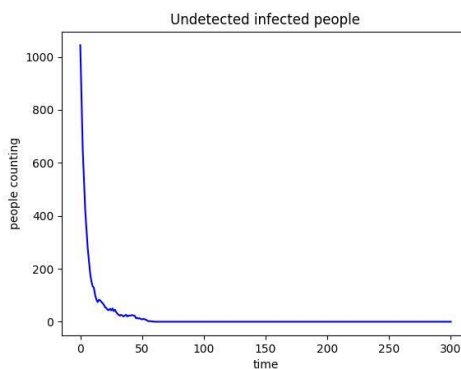


图 6-1-4 确诊人数

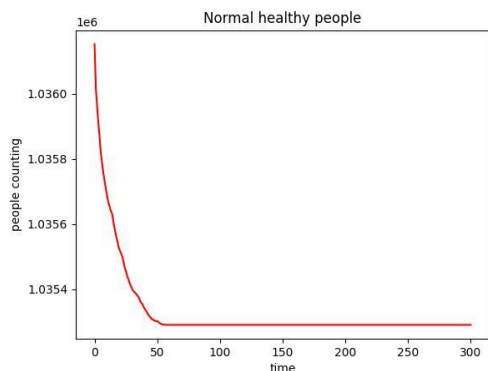


图 6-1-5 易感染者人数

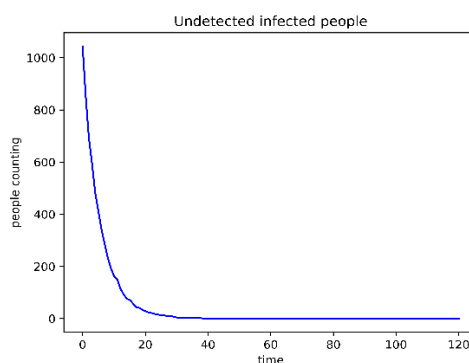


图 6-1-6 携带者人数

在对整个社会的代价和收益方面，关于物理隔离和免疫屏障的比较。参照 2020 年上半年全国施行封闭性管理物理隔离时期，做出如下对比：

1、相同点：均需要国家财政的巨大支持，医疗设备的大量供应。均能对疫情防控起显著的积极作用。均能对疫情后期的复工复产起安全保障作用。

2、不同点：

物理隔离：。经济方面，对于第一、二、三产业的 GDP 增长产生了较大程度的负面影响，整体 GDP 呈小幅度下降。国家资助防疫资金上千亿，支持复工复产支出两万亿。疫情成为劣汰优胜的媒介，对全国各企业进行了一次“换血”。线上消费迅速发展，线下经营惨淡；失业人数剧增，人民生活水平受限。社会稳定方面，未来新冠疫情复发会是对国家和整个社会的又一次重创，也会再次考验群众的心理承受能力，扰乱民心、破坏社会稳定的威胁将持续存在。

免疫屏障：经济方面，国家支出 5600 亿左右成本，实现全国各行各业的正常运行，GDP 在较大幅度波动下基本保持正常增长趋势。社会稳定方面，形成免疫屏障可有效消除群众对于新冠肺炎疫情的焦虑心理。

## 6.2 问题二的模型建立与求解

按照题目要求，我们选取杭州为例。首先，我们使用了第一问所用的模型去模拟杭州在疫情期间的病毒传播情况。建立模型曲线如图 6-1-4 所示。使用差分法不难得出图像所在的拐点，初始情况下计算拐点在步长为 58 处。我们做了一系列的实验求出拐点出现的时刻所在的概率曲线如图所示：

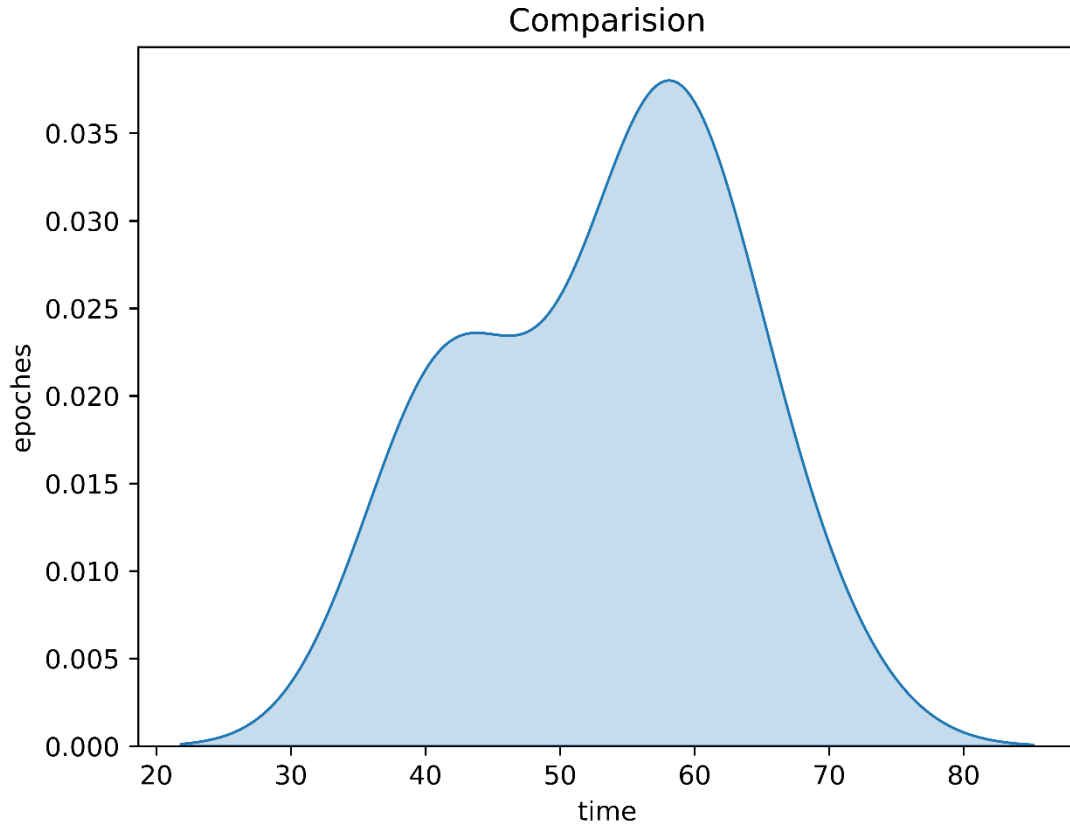


图 6-2-1

图像有两个峰值，在小峰值处病毒的作用较弱，图像在第 40 多轮已经达到拟合，需要达到群体免疫的轮数小；在第二处疫苗的作用较强，群众注射疫苗的几率较大，容易达成拟合且轮次较高，但达到拟合的概率比“安全”病毒的大些。这说明我们可以在接近 58 轮左右时去计算所需要的成本数。计算成本的公式如下：

$$W_{cost} = [t_0 \ t_1 \ t_2 \ t_3 \ t_4] \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \quad (5)$$



6.3 问题三的建立与求解

6.3.1 模型的建立

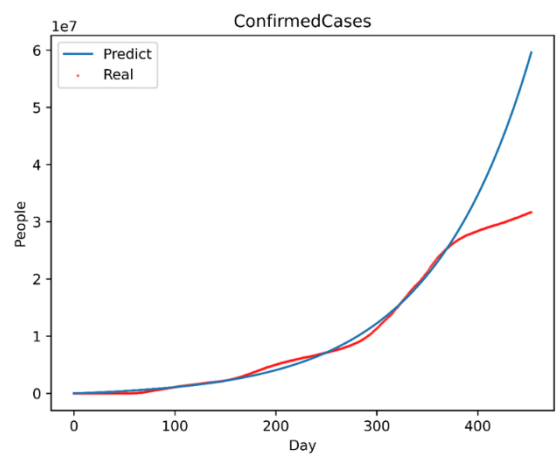


图 6-3-1 2020 年 1 月 1 日—2021 年 2 月 1 日  
美国累计确诊人数统计

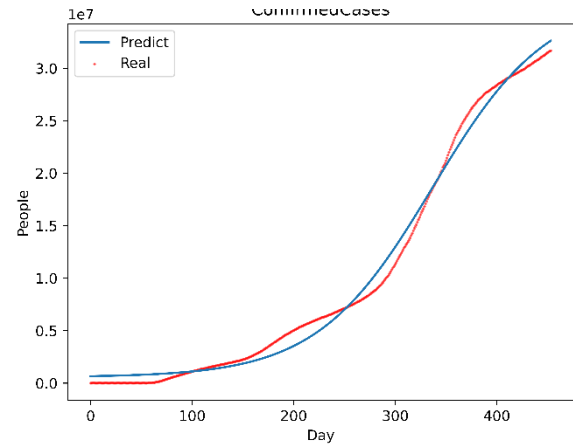


图 6-3-2 2020 年 1 月 1 日—2021 年 2 月 1 日



图 6-3-3 2020 年 1 月 1 日至 2021 年 2 月 1 日  
美国累计死亡人数统计与后期预测

根据 2020 年 1 月 1 日至 2021 年 5 月 1 日美国的疫情数据，利用回归分析得到公式 (6) 利用公式 (6) 对实际累计确诊人数与时间的函数关系进行拟合，并多次运用回归分析对 2021 年 2 月 1 日接种疫苗后的确诊人数增长趋势进行拟合。

$$N_{Case} = \frac{k_0}{1 + e^{-(k_1 x + k_2)}} + k_3 \quad (6)$$

如图 5-3，红色曲线代表实际确诊人数，由图可知 2021 年 2 月 1 日后累计确诊人数增幅显著降低；用蓝色曲线拟合累计确诊人数随时间的变化情况，公式 (6) 的各参数值为：

。

$$k_0 = 1.6180 \times 10^{10}$$
$$k_1 = 1.0100 \times 10^{-2}$$
$$k_2 = -10.1700$$
$$k_3 = -6.200 \times 10^5$$

2021 年 2 月 1 日后红蓝曲线的差异说明了疫苗接种率对美国疫情的积极影响。

如图 5-4，利用公式(6)回归分析得到参数为：

$$k_0 = 3.61 \times 10^7$$
$$k_1 = 1.7631 \times 10^{-2}$$
$$k_2 = -5.9376$$
$$k_3 = 5.6348 \times 10^5$$

得到对美国未来疫情中确诊人数的预测。

再次回归分析，利用公式 (6)，得到新的参数，结果如图 5-5 所示。

$k_0 = 6.2249 \times 10^7$
$k_1 = 3.7087 \times 10^{-3}$
$k_2 = -5.9322$
$k_3 = -1.4973 \times 10^5$

## 七、模型分析与检验

### 7.1 问题一的灵敏度分析

检验元胞自动机模型的灵敏度时，我们选取一组基准概率作为初始状态，即初始携带者人数占比  $P_1=0.1\%$ ，初始确诊者比例  $P_2=1.1\%$ ，初始疫苗接种者比例  $P_3=0.1\%$ ，无初始死亡者，其余均为初始易感人群。并设出各类状态转化的具体概率，如下表：

后状态 前状态	0	1	2	3	4
0	$P_{00}=0$	$P_{01}=0.008$	0	$P_{03}=0.002$	$P_{04}=0.005$
1	0	$P_{11}=0.695$	$P_{12}=0.007$	0	$P_{14}=0.005$
2	0	0	$P_{22}=0.79$	$P_{23}=0.1$	$P_{24}=0$
3	0	0	0	1	0
4	0	0	0	0	1

表 7-1 基准情况

每一次改变一个参数，并观察达到群体免疫时总计需要的步数。统计如下表：

改变参数	当前量	达到群体免疫时总计步数
------	-----	-------------

P <sub>01</sub>	0.0060	40
P <sub>01</sub>	0.0090	67
P <sub>01</sub>	0.0070	43
P <sub>01</sub>	0.0075	58
P <sub>01</sub>	0.0065	42
P <sub>03</sub>	0.0030	58
P <sub>03</sub>	0.0010	59
P <sub>03</sub>	0	59
P <sub>03</sub>	0.0050	54
P <sub>2</sub>	0.001	41
P <sub>2</sub>	0.02	崩溃
P <sub>3</sub>	0.002	41
P <sub>3</sub>	0	33

表 7-2 灵敏度实验

经过十余次变换参数，可以发现每次在 $\pm 100\%$ 范围内改变概率大小时，达到群体免疫的总计步数多数集中在 58 步，少部分多于或少于 58 步。由此说明该模型的稳定性高，在参数变化的一定范围内能保持稳定，也不至于迟钝，对于参数变化有相应的反应。

## 7.2 问题三的模型检验

问题三要求对美国疫情做出疫苗接种对疫情的影响和疫情的未来趋向预测。人口的拟合程度高低是对预测模型准确度的主要标准。对美国未来疫情趋势的数学模型进行检验，发现美国的累计确诊人数统计及预测图的回归系数  $\text{ConfirmedCases } r^2=0.9973$ ，累计死亡人数统计及预测图的回归系数是  $\text{ConfirmedDeaths } r^2=0.9846$ 。显然我队所选模型拟合度极高，能以较高的精准度反映美国未来的疫情走向。

## 八、模型评价与推广

在本次建模过程中，我们所用到的元胞自动机本身就是做局部仿真和局部联系的利器。然而在模型中描述的有限的状态，却在与实际参数最小再生数相撞后爆发出无限的可能。一旦参数全部确定，我们的模型就稳定地输出，曲线收敛的步长相对稳定，曲线的趋势在一定范围内修改参数后大体不会改变。很好地满足了我们模拟病毒传播的需求，并且在求解出参数间的关系后愈发逼近实际情况。

元胞自动机的推广应用有很多，它功能简单，除了病毒的模拟，还可以将它应用在森林火灾模拟，真菌生长模拟。只要适应情况地变化规则，就能模拟想要的数学模型。

元胞自动机缺点在于：具体的参数无法直接找出，不同参数进行模拟时调参具有随机性。

## 九、参考文献

- [1]张玥:《中国抗疫账本: 千亿防控, 两万亿扶持, 八万亿计划》, [中国抗疫账本: 千亿防控, 两万亿扶持, 八万亿计划 | 南方周末 \(infzm.com\)](#) , 2021 年 5 月 4 日
- [2]樊非: 新型冠状病毒疫情的建模与分析[D] 2020