原文地址：<https://zhuanlan.zhihu.com/p/104379096>

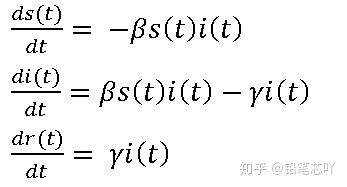
2019年底至今，中国爆发了新型冠状病毒肺炎，疫情席卷全国，是继非典（sars病毒）后又一全国性的病毒性肺炎传播。SIR模型（Susceptible Infected Recovered Model）是常用的传染病模型，用此模型对本次的疫情发展趋势进行分析具有一定的参考价值。

模型假设暂不考虑人口的出生与死亡，人口总量视为定值，治愈后个体具有对病毒的免疫能力；设单位时间内的易感人群表示为s(t),感染人群为i(t),恢复人群为r(t);有如下推断：

1. 单位时间内感染人群i(t)所能感染的人数与易感人群s(t)成正比，比例系数为β，单位时间感染的人数为βs(t)i(t);

2. 单位时间内恢复人群与感染人群i(t)成正比，比例系数为γ，单位时间恢复的病人数为γi(t).

单位时间内各变量的变化情况即为变量的增长率，故上述推断可描述为如下微分方程的形式：



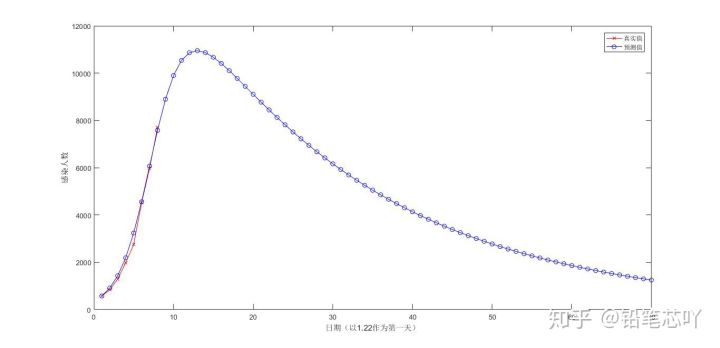
要得到该方程的解析解十分困难，可借助matlab解得其数值解进行分析和预测，预测的关键在于感染系数β与恢复系数γ的取值。可搜集1.22起至今日的数据进行估计。

同样地，考虑到对已诊断人群的隔离、春节假期间武汉市进行封锁，且群众往往应要求宅在家里；当下的环境并不是一个感染人群和普通群众充分混合的环境；故不能使用全国人口数作为人口总量进行预测，需要估计易感人群初值s(0).

由推断2不难看出恢复系数γ可由感染人群和恢复人群的数据计算得到，而感染人群和恢复人群是可以直接观测的；但当下所能调查的数据较为有限，1.22日至今甚至没有达到正常病毒性肺炎的恢复周期，且恢复人数截至今日也只有百余例，感染人数却在快速上升；用现有数据计算出的恢复系数过低，恢复周期长达150天，与现实严重不符。查阅资料发现一般病毒性肺炎恢复周期为2-3周，非典的恢复周期约为22天，故将此次病毒性肺炎恢复周期估计为25天，恢复系数γ≈1/25=0.04.

对于易感人群初值s(0)和感染系数β的估计采用最小二乘估计的方法，最小二乘估计的基本思路是先估计s(0)和β的取值区间，然后寻找区间内计算得到预测值与真实值残差平方和最小的一组值作为估计值进行预测。本方案遍历的s(0)区间为[5000,30000].

预测结果如下，感染系数β≈3.57\*10^-5.



从图中可以看出，近五天就是拐点，感染人数增长率放缓；2月3日左右到达峰值，有可能破万，达到11000人左右。

本预测仅供参考，存在的局限性在于：

1. 未考虑新型冠状病毒肺炎的潜伏期，与非典型肺炎相比，本次蔓延的病毒有更长的潜伏期，不易察觉；

2. 未能根据已有数据很好地估计恢复系数，用更详实的数据或其他更好的方法估计恢复系数应是改进预测的一大思路；

3.真实的感染人数不等同于当日的确诊人数，当日的疑似症状病人数应有部分感染，且不可避免地存在漏报甚至瞒报等情况，实际的感染人数是不可观测的。

数据来源：央视新闻微博

参考文献：徐宝春. 山东大学. 基于SIR模型的SARS传染病研究

matlab代码如下：

forecast.m

clear all;clc;clf;

i = [571 830 1287 1975 2744 4515 5974 7711];r = [28 34 38 49 51 60 103 124];%i(t) r(t)初值

%估计恢复系数

% for n = 1:6

% r0(n) = (r(n+1) - r(n))./i(n);

% end

% b = mean(r0);

b = 0.04;

%用最小二乘法估计s(t)初值和感染系数a

s0 = 1500:10:30000;

for n = 1:length(s0)

s1 = s0(n) + i(1) - i(2) - r(2);

a = (s0(n) - s1)./(s0(n).\*i(1));

t = 1:8;

[T,Y] = ode45('sir',t,[s0(n),i(1),r(1),a,b]);

for p = 1:8

SE(p) = (Y(p,2) - i(p)).^2;

end

SEE(n) = sum(SE); %此时的残差平方和

end

[v,address] = min(SEE);

s = s0(address); %残差平方和最小的s(t)初值

s1 = s + i(1) - i(2) - r(2);

a = (s - s1)./(s.\*i(1)); %残差平方和最小的感染系数a

tspan = 1:70;

[T,Y] = ode45('sir',tspan,[s,i(1),r(1),a,b]);

%作图

figure(1);

t = 1:8;

plot(t,i,'rx-');

hold on;

y = Y(:,2);

y = y';

plot(T,y,'bo-');

legend('真实值','预测值');

xlabel('日期（以1.22作为第一天）');ylabel('感染人数');

sir.m

function [ y ] = sir(t,x)

%描述sir模型的微分方程

%数组x前三个值分别为s i r初值，后两个值存储感染系数a 恢复系数b

y=[-x(4)\*x(1)\*x(2),x(4)\*x(1)\*x(2)-x(5)\*x(2),x(5)\*x(2),0,0]'; %s增长率，i增长率，r增长率

end