

# 传染病早期传播模型: 分支过程

## 生物数学课程报告

柴昊

上海数学中心

2022.12.5

# 目录

- 1 传染病早期传播模型
- 2 基本传染数和灭绝概率
- 3 数据分析与讨论
- 4 致谢

## 传染病早期传播模型

# 模型介绍

## 文章来源:

Oliveira, G. Early epidemic spread, percolation and Covid-19. J. Math. Biol. 81, 1143–1168 (2020). <https://doi.org/10.1007/s00285-020-01539-1>

## 主要内容:

假设一类传染病可以在人类社会传播, 利用分支过程对传染病早期传播 (感染者数量较少) 进行建模, 假设早期传播过程中存在多种传播途径 (感染概率不同), 利用生成函数计算传染病的基本传染数  $R_0$  以及传染病被控制住的概率  $P_\infty$ 。最后根据早期新型冠状病毒的数据 (COVID-19) 估算了一些国家 (德国、巴西等) 的模型参数。

- 假设感染者独立/非独立地发生  $n$  种传播方式  $T_1, \dots, T_n$
- 每种传播方式随机地接触到  $k_1, \dots, k_n$  个易感染者
- 每种传播方式有  $T_1, \dots, T_n$  的概率造成传染从而产生  $m_1, \dots, m_n$  个感染者, 接触过程和感染过程, 感染过程之间是彼此独立的
- 假设人类社会对该传染病进行防控使得任何一个感染者有  $Q$  的概率接受治疗/隔离从而治愈/失去传染能力

从一个零号感染者  $X_0$  或者一个随机感染者  $\tilde{X}_0$  进行跟踪, 试图找出传染病传播  $m$  代际后感染人数  $X_m$  的概率分布情况, 传染病在有限时间内可以被控制的概率  $P_\infty$  (灭绝概率), 以及平均单个感染者传染的人数  $R_0$  (基本传染数)。

















## 分支过程模型的有关参数

### 传播过程的随机性质

继续分析随机变量  $m$ 。首先考虑  $(T^{(1)}, \dots, T^{(n)})$  的联合分布, 记  $\mathbb{P}_i$  是随机变量  $T^{(i)}$  的边际分布。考虑以下两种情况

- 传播相互独立情况 (例如考虑一个确定感染的病人进行的传播过程), 这时有联合分布的概率函数是

$$\mathbb{P}(T^{(i)} = k_i, i = 1, \dots, n) := \prod_{i=1}^{i=n} \mathbb{P}(T^{(i)} = k_i) = \prod_{i=1}^{i=n} \mathbb{P}_i(k_i)$$

- 传播存在依赖情况 (对接触人数进行加权), 这时的联合分布是

$$\mathbb{P}(T^{(i)} = k_i, i = 1, \dots, n) := \frac{\sum_{i=1}^n (k_i + 1) \mathbb{P}_i(k_i + 1) \prod_{j \neq i} \mathbb{P}_j(k_j)}{\mathbb{E} T^{(1)} + \dots + \mathbb{E} T^{(n)}}$$





## 基本传染数和灭绝概率

# Galton-Watson 过程的灭绝概率

## 定义

上述考虑的分支过程最早是 Galton 在 19 世纪研究英格兰地区中世纪姓氏的消亡时提出的。而后又被 Watson 等人重新发现。一个自然的问题就是一个姓氏有多大的可能逐渐消失，对应于传染病模型或者基因遗传模型就是计算如下概率

$$\mathbb{P}(\lim_{n \rightarrow +\infty} S_n = 0) = ?$$

根据这个分支过程的性质

(如果第  $n$  代灭绝则后续代际均灭绝,  $S_n = 0 \Rightarrow S_{n+k} = 0, \forall k \in \mathbb{N}$ )

可以知道

$$\mathbb{P}(\lim_{n \rightarrow +\infty} S_n = 0) = \mathbb{P}(\exists N > 0, S_n = 0, \forall n \geq N) = \lim_{n \rightarrow +\infty} \mathbb{P}(S_n = 0) = \mathbf{P}_\infty$$

这个概率  $\mathbf{P}_\infty$  记作灭绝概率。





# Galton-Watson 过程的灭绝概率

图像说明

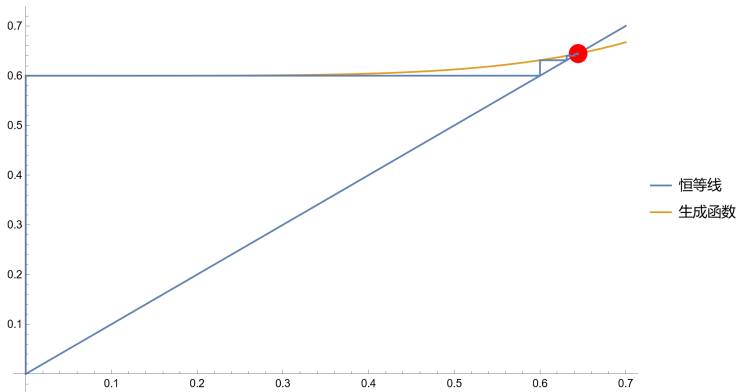


图: 蛛网工作法 (从  $x = 0$  出发)

# Galton-Watson 过程的灭绝概率

图像说明

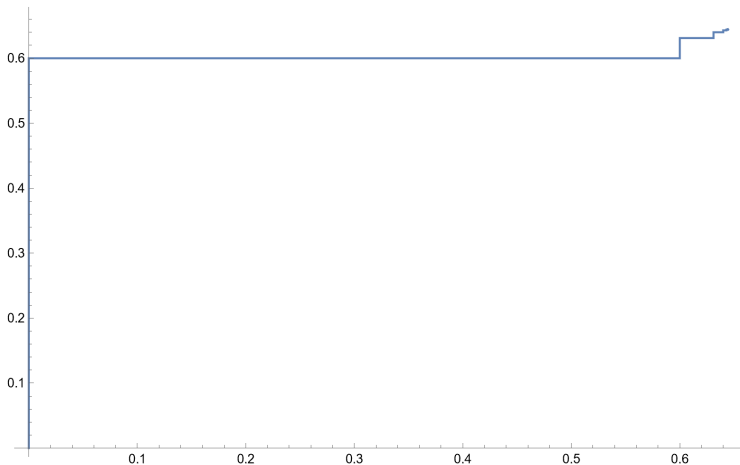


图: 迭代折线不会超过最小不动点

## 计算基本传染数 $R_0$

### 定义

考虑单个感染者传染人数的随机变量  $m = m_1 + \cdots + m_n$  的期望  $R_0 = \mathbb{E}m$ 。根据生成函数的理论可知事实上考虑了防控、社交隔离等抑制因素下某种传染病的基本传染数  $R_0$  是生成函数在  $x = 1$  处的导数

$$R_0 = \frac{d}{dx} \tilde{Z}(1), \text{ or } R_0 = Z'(1), \tilde{Z}'(1)$$

根据之前对生成函数的计算, 可以得到分别取  $z, \tilde{z}$  计算时有

$$R_0 = (1 - Q) \sum_{i=1}^{i=n} T_i \frac{\partial G}{\partial x_i} \Big|_{x_1=\dots=x_n=1}$$

$$R_0 = (1 - Q) \sum_{i=1}^{i=n} T_i \frac{\partial \tilde{G}}{\partial x_i} \Big|_{x_1=\dots=x_n=1}$$

## 计算基本传染数 $R_0$

### 利用生成函数关系计算

利用  $G$  与  $\tilde{G}$  存在如下关系

$$\tilde{G}(x_1, \dots, x_n) = \frac{\sum_{i=1}^n G_i''(x_i) \prod_{j \neq i} G_j(x_j)}{G_i(1)}$$

### 利用生成函数与矩的联系不难计算出偏导数

$$\begin{aligned} \frac{\partial \widetilde{G}}{\partial x_i} \Big|_{x_1=\dots=x_n=1} &= \frac{G''_i(1) + G'_i(1) \sum_{j \neq i} G'_j(1)}{\sum_{j=1}^{j=n} G'_j(1)} \\ &= \frac{\mathbb{E} T^{(i)^2} - \mathbb{E} T^{(i)} + \mathbb{E} T^{(i)} \sum_{j \neq i} \mathbb{E} T^{(j)}}{\mathbb{E} T^{(1)} + \dots + \mathbb{E} T^{(n)}} \\ &= \frac{\mathbf{Var} T^{(i)}}{\mathbb{E} T^{(1)} + \dots + \mathbb{E} T^{(n)}} + \mathbb{E} T^{(i)} \left(1 - \frac{1}{\mathbb{E} T^{(1)} + \dots + \mathbb{E} T^{(n)}}\right) \end{aligned}$$

从而可以用  $T^{(i)}, T_i$  计算基本传染数  $R_0$

$$R_0 = (1 - Q) \sum_{i=1}^{i=n} T_i \left( \frac{\mathbf{Var} T^{(i)}}{\mathbb{E} T^{(1)} + \dots + \mathbb{E} T^{(n)}} + \mathbb{E} T^{(i)} \left( 1 - \frac{1}{\mathbb{E} T^{(1)} + \dots + \mathbb{E} T^{(n)}} \right) \right)$$

## $R_0$ 与预测传染病发展

之前提到了  $R_0$  的数值意义可以理解为单个感染者平均能够传染的人数，这个数据在传染病学中常常作为衡量传染病传播能力的重要参数。理论上有如下结论。

### $R_0$ 判据

如果上述传染病模型的  $R_0 < 1$  则传染病几乎必然会消失 (灭绝概率为  $P_\infty = 1$ , 一定会被控制); 如果  $R_0 > 1$  则存在非零的灭绝概率  $1 > P_\infty > 0$ 。

这个结论可以简单说明如下, 由于生成函数  $Z(x), \tilde{Z}(x)$  都是下凸的, 那么根据中值定理可知

$$|Z(x_1) - Z(x_2)| \leq |Z'(1)||x_1 - x_2| = R_0|x_1 - x_2|, \quad x_1, x_2 \in [0, 1]$$

如果  $R_0 < 1$  可知  $Z, \tilde{Z}$  是收缩映射, 根据 Banach 不动点定理可知其在  $[0, 1]$  上有唯一的不动点  $x_0 = 1$  对应了唯一的灭绝概率是  $P_\infty = 1$ 。另一方面如果  $R_0 > 1$  而且  $Z(0), \tilde{Z}(0) > 0$  对  $Z(x) - x$  或者  $\tilde{Z}(x) - x$  应用费马中值定理可知在  $(0, 1)$  区间上有不动点  $x_{00}$  对应了一个非零的灭绝概率。

# $R_0$ 与预测传染病发展

图像说明

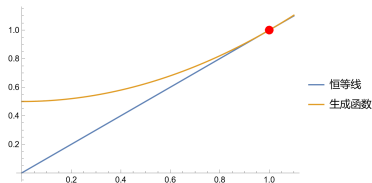


图: 基本传染数小于一

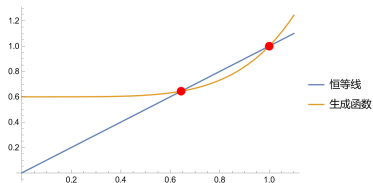


图: 基本传染数大于一

## 最小隔离率

如果可以人为控制隔离政策和监测方式即增加  $Q$  值, 可以预见到这样做会使得传染病演化为流行病的可能性减小。自然要问  $Q$  至少增大到多少才能使得传染病不可能一直传播下去, 也就是采取多强有力的防控政策可以控制住传染病的传播。这个最小值就是最小隔离率, 记作  $Q_{min}$ 。

根据上面的推导可以得到一个下界

$$Q_{min} > 1 - \frac{1}{R_0^{\text{free}}}$$

这里的  $R_0^{\text{free}}$  是指不进行任何政策干预的情况下 ( $Q = 0$ ) 上述模型的基本传染数

$R_0^{\text{free}} = R_0(0) = \frac{dZ^{\text{free}}}{dx}|_{x=1}$ 。这里的  $Z^{\text{free}}$  (或取  $\widetilde{Z^{\text{free}}}$ ) 是将参数  $Q$  设置为零的生成函数。

这是因为无论对于  $Z$  还是  $\widetilde{Z}$  而言都有

$$Z(x) = Q + (1 - Q)Z^{\text{free}}(x), \quad \widetilde{Z}(x) = Q + (1 - Q)\widetilde{Z^{\text{free}}}(x)$$

那么取  $x = 1$  处的导数就有

$$R_0 = (1 - Q)R_0^{\text{free}}$$

为了完全控制传染病流行需要  $R_0 < 1$  所以代入上式得到  $Q_{min} > 1 - \frac{1}{R_0^{\text{free}}}$ 。



## 数据分析与讨论

# 实际建模

## 考虑两类接触方式

为了简洁这里选取论文中的一个例子，套用之前两节讲解的理论进行建模。  
对于社会上的可能被感染/可能有传染力的人，假设其在日常生活中有两类明显不同的接触类型，一种是和熟人（例如亲人朋友）的接触，这种接触的人数上限可以假设是有限的（假设普通人的朋友不可能无穷多），而且接触后染病的概率也应该相应高一些，这样有

$$G_1(x) = p_0 + p_1 x + \cdots + p_n x^n = \sum_{s=0}^{s=n} \mathbb{P}(k_1 = s) x^s$$

是一个多项式分布（生成函数是一个多项式，分布是有限值分布）。  
第二种接触是和陌生人的接触（例如乘坐公共交通进出人员聚集的公共场所），这种接触可以近似成一个无记忆的等待队列（离散泊松过程），根据概率论的知识和课程中的讲授可比较合理地将其近似为泊松分布  $\text{Poisson}(\lambda)$ ，于是生成函数如下

$$G_2(x) = \sum_{s=0}^{+\infty} \frac{\lambda^s}{s!} e^{-\lambda} x^s = e^{-\lambda} \sum_{s=0}^{+\infty} \frac{(\lambda x)^s}{s!} = e^{\lambda(x-1)} = \exp(\lambda x - \lambda)$$

根据上面假设可以得到  $\tilde{G}, \tilde{Z}, R_0$  等关键参数的表达式。

# 实际建模

## 参数的估计和计算

在该论文中详细讨论了以下几种传播模式的计算

- 有限分布/泊松分布
- 泊松分布/泊松分布
- 带有隔离政策和防控意识  $Q, f_1, f_2$  的泊松分布/泊松分布
- 带有隔离政策和防控意识  $Q, f_1, f_2$  的多项式分布/泊松分布

由于上述讨论都是繁琐的计算，限于报告篇幅就不详细说明了。

论文中作者取  $G_1$  是一个五次多项式,  $G_2$  是一个带有隔离概率的泊松分布母函数即

$$G_1(x) = p_0 + \cdots + p_5 x^5, \quad G_2(x) = f + (1 - f)e^{-N_2(1-x)}$$

从而

$$\tilde{Z}(x) = Q + (1 - Q) \frac{G_1'(T_1 x + (1 - T_1)) G_2(T_2 x + (1 - T_2)) + N_2 G_1(T_1 x + (1 - T_1)) e^{-N_2 T_2 (1 - x)}}{G_1'(1) + N_2}$$

$$R_0 = Q + (1 - Q) \frac{T_1 G_1''(1) + (1-f)N_1 N_2 (T_1 + T_2) + (1-f)N_2^2 T_2}{G_1'(1) + N_2}$$

应用了 2020 年初期德国, 西班牙, 葡萄牙, 巴西等国新冠病毒 (COVID-19) 感染者的数据估计了必要参数如  $Q, T_{i,f}, p_i$  等, 然后计算了  $R_0$  和  $P_\infty$ 。如下表。

国家	$R_0$	$\mathbf{P}_\infty$	$Z(\mathbf{P}_\infty)$
德国	2.81	0.0852	0.0855
西班牙	3.44	0.039	0.039
葡萄牙	3.4	0.04	0.04
巴西	3.82	0.026	0.026

# 一些讨论

为什么在传染病流行后期，分支过程的建模可能不再准确？需要考虑偏微分方程模型如 SIR 模型。

# 一些讨论

为什么在传染病流行后期，分支过程的建模可能不再准确？需要考虑偏微分方程模型如 SIR 模型。

是否可以改善这个模型对传染病发展进行预测，进而根据预测结果对模型进行修正 (例如修正参数)？

## 致谢

报告结束 感谢聆听