

数据驱动下新冠肺炎基本再生数的计算方法

杨俊元¹,张烁¹,王国强¹,李学志²

(1.山西大学 复杂系统研究所,太原 030006;2.河南师范大学 数学与信息科学学院,河南 新乡 453007)

摘要:2019 年突如其来的新冠肺炎疫情给我国经济社会发展和人民生活带来严重影响,如何科学合理计算新冠肺炎传播的基本再生数对疫情的控制尤为重要。目前,已有的计算新冠肺炎基本再生数方法大都基于数据辨识模型参数进而算出其值,给出一种利用数据结合模型直接计算新冠肺炎传播基本再生数的新方法。利用该方法及峰值数据可以计算基于 SIR 模型、SEIR 模型及 SEIAR 模型全国、湖北及广东新冠肺炎传播的基本再生数,该方法能反映基本再生数和新冠肺炎传播相关数据的直接关系。

关键词:基本再生数;SIR 模型;SEIR 模型;峰值

中图分类号:O29

文献标志码:A

自 2019 年 12 月首次报道新冠肺炎(COVID-19)感染以来,目前全国有 8 万多人确诊,3 千多人死亡^[1-2]。虽然经过全国人民的不懈努力,取得了阶段性胜利,但也付出了极大的代价,目前所有大、中小学校仍没有开学,部分企业还没复工,全球疫情形式仍不乐观,韩国、日本、伊朗、意大利等国新冠肺炎病例数仍在持续上升,要想彻底消除该病道路依旧漫长,如何有效控制疫情的传播和蔓延是所有国家共同关心和亟待解决的问题^[3]。

数学模型能很好地揭示传染病传播的主要因素和规律,鉴于新冠肺炎(COVID-19)的传播机理,许多学者利用 SIR、SEIR 和 SEIAR 等仓室模型来研究新冠肺炎在中国的传播、预测及控制措施评价^[4-7],为疫情的控制提供了理论和科学评价依据。

基本再生数是指一个病人在平均患病周期内传染的病人数^[8],基本再生数的大小和疾病的流行与灭绝有着必然的联系,一般来说,基本再生数大于 1,疾病暴发;否则疾病消亡。新型冠状病毒肺炎(COVID-19)是一种突发的急性传染病,其初期的增长速度和基本再生数有着必然的联系,如何精确估算新型冠状病毒肺炎(COVID-19)的基本再生数是众多学者共同关心的问题,目前普遍的方法是通过数据拟合模型中的参数,进而估计基本再生数,这种方法依赖于数据的数量和初值选取的方法,得到的基本再生数也有很大的差异,本文利用国家公布有关新冠肺炎(COVID-19)疫情数据,如累计确诊病例、累计治愈人数、潜伏期、平均染病周期、初值及峰值直接计算新冠肺炎的基本再生数,能克服参数估计带来的误差,该方法有效分析疾病传播及流行的新方法。

1 主要结果

1.1 SIR 模型

新冠肺炎(COVID-19)是一种急性突发传染病,具有流行时间较短,通过合理治疗可以康复等特点,故可以将总人口 N 分成 3 种不同的状态:易感类、染病类和康复类,分别用 $S(t)$, $I(t)$ 和 $R(t)$ 表示,染病类以

收稿日期:2020-03-07;修回日期:2020-03-09.

基金项目:国家自然科学基金(61573016;61873154;11771017);山西省新型冠状病毒(COVID-19)预防和控制项目(202003D31011/GZ).

作者简介:杨俊元(1981—),男,山西朔州人,山西大学教授,研究方向为生物数学,复杂网络.

通信作者:李学志,河南师范大学教授,博士生导师,E-mail:xzli66@126.com.

速度 $\beta \frac{I}{N}$ 感染易感类; 染病类经过 $1/\gamma$ 天的治疗转化成康复类. 这个过程可以用下述方程表示.

$$\begin{cases} \frac{dS}{dt} = -\beta \frac{SI}{N}, \\ \frac{dI}{dt} = \beta \frac{SI}{N} - \gamma I, \\ \frac{dR}{dt} = \gamma I, \end{cases} \quad (1)$$

国家健康委员会于 2020 年 1 月 11 日开始正式报道新冠肺炎 (COVID-19) 病例数, 因此可以假设 1 月 10 日为初始时刻 0. 假设方程 (1) 带有如下初值

$$S(0) = S_0, I(0) = I_0, R(0) = R_0.$$

由文献 [8] 的方法可以算出模型 (1) 的基本再生数 $\mathcal{R}_0 = \frac{\beta}{\gamma}$. 假设新冠肺炎 (COVID-19) 在 T 时刻达到峰值, 则

$\frac{dI}{dt} \Big|_{t=T} = 0$ 且 $I(T) > 0$. 由模型 (1) 第 2 个方程两边消去 $I(T)$ 得:

$$S(T) = \frac{N}{\mathcal{R}_0}. \quad (2)$$

对模型 (1) 的第 2 个方程从 0 到 T 积分得:

$$S_0 + I_0 - S(T) - I(T) = \gamma \int_0^T I(t) dt. \quad (3)$$

利用常数变易法求解模型 (1) 的第 1 个方程得:

$$\int_0^T I(t) dt = \frac{N}{\beta} \ln \left(\frac{S_0}{S(T)} \right). \quad (4)$$

将 (4) 式代入 (3) 式得:

$$S_0 + I_0 - S(T) - I(T) = \frac{N}{\mathcal{R}_0} \ln \left(\frac{S_0}{S(T)} \right). \quad (5)$$

将 (2) 式代入 (5) 式得:

$$S_0 + I_0 - \frac{N}{\mathcal{R}_0} - I(T) = \frac{N}{\mathcal{R}_0} \ln \left(\frac{S_0 \mathcal{R}_0}{N} \right). \quad (6)$$

假设新冠肺炎在中国流行期间, 易感人群的初值近似取为 $S(0) = \alpha N$ (α 为高危人群的占比), 则有如下计算基本再生数的方法.

定理 1 假设初始时刻 $I_0 = 0$, 且染病峰值为 $I(T)$, 则 \mathcal{R}_0 是如下隐式方程的解

$$N \left(\alpha - \frac{1}{\mathcal{R}_0} \right) - I(T) = \frac{N \ln(\alpha \mathcal{R}_0)}{\mathcal{R}_0}. \quad (7)$$

或者已知累计出院数, 由下面定理计算新冠肺炎的基本再生数.

定理 2 假设初始时刻 $R_0 = 0$, 在峰值时刻累计出院人数为 $R(T)$, 则 \mathcal{R}_0 也可用如下公式计算

$$R(T) = \frac{N \ln(\alpha \mathcal{R}_0)}{\mathcal{R}_0}. \quad (8)$$

由国家疾病与健康委员会网站公布的数据^[9]和中国人口统计年鉴数据^[10], 可以统计得到全国新增新冠肺炎染病数、累计病例数、新增出院数及累计出院数. 进一步计算可得全国、湖北、广东分别于 2 月 17 日、2 月 18 日和 2 月 10 日达到峰值 59 884 例、52 554 例和 1 008 例 (见图 1). 将这些峰值代入 (7) 式计算出全国、湖北及广东新冠肺炎传播基本再生数 (见表 1).

在 SIR 模型框架下, 表 1 表明随着高危人群易感比例的上升, 基本再生数在逐渐减小. 在易感高危比例相同的条件下, 广东省的基本再生数要高于全国和湖北.

1.2 SEIR 模型

上一节讨论的模型忽略了疾病的潜伏期. 据报道, 新冠状肺炎的潜伏期 3~14 d, 平均潜伏期为 7 d. 鉴于新冠肺炎潜伏性, 许多学者引入一个潜伏仓室 E 表示 COVID-19 的潜伏状态. 易感个体被感染后, 先转化成潜伏类, 而后经过 $1/q$ 天潜伏期发病. 上述新冠肺炎传播过程可以用如下仓室模型表达.

$$\begin{cases} \frac{dS}{dt} = -\beta \frac{SI}{N}, \\ \frac{dE}{dt} = \beta \frac{SI}{N} - qE, \\ \frac{dI}{dt} = qE - \gamma I, \\ \frac{dR}{dt} = \gamma I, \end{cases} \quad (9)$$

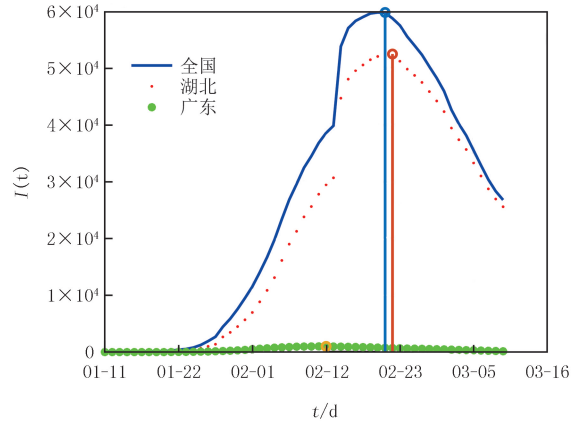


图1 新冠肺炎在中国湖北及广东流行趋势图

Fig.1 The trends of COVID-19 infection in China, Hubei and Guangdong

表 1 新冠肺炎在中国、湖北及广东流行趋势图

Tab.1 The trends of COVID-19 infection in China, Hubei and Guangdong

区域	$\alpha=0.2$	$\alpha=0.4$	$\alpha=0.6$	$\alpha=0.8$	$\alpha=1$
全国	$\mathcal{R}_0 = 4.896\ 0$	$\mathcal{R}_0 = 2.463\ 0$	$\mathcal{R}_0 = 1.646\ 5$	$\mathcal{R}_0 = 1.236\ 9$	$\mathcal{R}_0 = 0.990\ 6$
湖北	$\mathcal{R}_0 = 4.563\ 2$	$\mathcal{R}_0 = 2.342\ 2$	$\mathcal{R}_0 = 1.579\ 9$	$\mathcal{R}_0 = 1.193\ 3$	$\mathcal{R}_0 = 0.959\ 3$
广东	$\mathcal{R}_0 = 4.953\ 2$	$\mathcal{R}_0 = 2.483\ 4$	$\mathcal{R}_0 = 1.657\ 6$	$\mathcal{R}_0 = 1.244\ 1$	$\mathcal{R}_0 = 0.995\ 8$

假设模型(9)的初值为 $S(0) = S_0, E(0) = E_0, I(0) = I_0, R(0) = R_0$. 同样利用文献[8]算出模型(9)基本再生数表达式为 $\mathcal{R}_0 = \frac{\beta}{\gamma}$. 注意到, 虽然模型(9)包含潜伏仓室但基本再生数的表达式与潜伏期 $1/q$ 无关. 为了方便, 仍用 T 表示染病个体达到峰值的时间. 利用第一节类似的方法, 得到

$$S_0 + E_0 + I_0 - S(T) - E(T) - I(T) = \frac{N}{\mathcal{R}_0} \ln \frac{S_0}{S(T)}. \quad (10)$$

这里 $S(T)$ 和 $E(T)$ 表示在峰值时刻易感者的数量和潜伏者的数量. 同理, 染病者在达到峰值时 $\dot{I}|_{t=T} = 0$. 那么由模型(9)的第 3 个方程可得 $E(T) = \frac{\gamma}{q} I(T)$. 将其带入到(10)式得如下定理.

定理 3 假设初始时刻 $I_0 = E_0 = 0$, 染病峰值为 $I(T)$ 和累计出院数为 $R(T)$, 则 \mathcal{R}_0 是如下隐式方程的

$$\text{解 } N(\alpha - 1) + R(T) = \ln \left(\frac{\alpha N}{N - \frac{\gamma}{q} I(T) - I(T) - R(T)} \right) \frac{N}{\mathcal{R}_0}.$$

注意到 $1/q$ 表示潜伏期, $1/\gamma$ 表示病程, 则由文献[7]得新冠肺炎的平均潜伏期为 7 d, 平均治疗时间为 10 d. 故 $q=1/7, \gamma=1/10$. 因此, 利用上面的数据和(11)式得出全国、湖北及广东省新冠肺炎传播的基本再生数(见表 2).

表 2 全国、湖北及广东新冠肺炎传播基本再生数

Tab.2 Basic reproduction number of COVID-19 in China, Hubei and Guangdong

区域	$\alpha=0.2$	$\alpha=0.4$	$\alpha=0.6$	$\alpha=0.8$	$\alpha=1$
全国	$\mathcal{R}_0 = 2.011\ 7$	$\mathcal{R}_0 = 1.527\ 0$	$\mathcal{R}_0 = 1.276\ 9$	$\mathcal{R}_0 = 1.115\ 3$	$\mathcal{R}_0 = 9.110\ 9$
湖北	$\mathcal{R}_0 = 2.010\ 1$	$\mathcal{R}_0 = 1.524\ 8$	$\mathcal{R}_0 = 1.273\ 4$	$\mathcal{R}_0 = 1.118\ 2$	$\mathcal{R}_0 = 10.796\ 7$
广东	$\mathcal{R}_0 = 2.011\ 8$	$\mathcal{R}_0 = 1.527\ 1$	$\mathcal{R}_0 = 1.277\ 0$	$\mathcal{R}_0 = 1.115\ 6$	$\mathcal{R}_0 = 12.983\ 3$

在 SEIR 模型框架下, 基本再生数与高危易感比例没有严格的单调关系, 如果易感人群 100% 暴露, 基本

再生数维持在一个比较高的水平.在易感比例是 80%的情况下,湖北省的基本再生数略高于全国和广东.

1.3 SEIAR 模型

另外,有报道认为无症状的新冠肺炎(COVID-19)病人可以传播新型冠状病毒.因此,引入一个仓室 A 来表达无症状的状态,且无症状的个体可以传染易感个体.则含无症状感染状态的新冠肺炎演化过程可表述为

$$\begin{cases} \frac{dS}{dt} = -\beta \frac{SI}{N} - \beta_1 \frac{SA}{N}, \\ \frac{dE}{dt} = \beta \frac{SI}{N} + \beta_1 \frac{SA}{N} - qE, \\ \frac{dI}{dt} = pqE - \gamma I, \\ \frac{dA}{dt} = (1-p)qE - \gamma_1 A, \\ \frac{dR}{dt} = \gamma I + \gamma_1 A. \end{cases} \quad (11)$$

这里 β_1 表示无症状个体传染易感者的速率, p 表示潜伏个体转化成染病个体的概率, $(1-p)$ 表示潜伏个体转化成无症状个体的概率, γ_1 表示无症状个体康复的概率.假设模型具有初值

$$S(0) = S_0, E(0) = E_0, I(0) = I_0, A(0) = A_0, R(0) = R_0.$$

利用文献[8]得出模型(11)基本再生数的表达式为 $\mathcal{R}_0 = \frac{p\beta}{\gamma} + \frac{(1-p)\beta_1}{\gamma_1}$.

由模型(11)的第 1 个和第 2 个方程得

$$\frac{\beta}{N} \int_0^T I(t) dt + \frac{\beta_1}{N} \int_0^T A(t) dt = \ln \left(\frac{S_0}{S(T)} \right), \quad (12)$$

$$S_0 + E_0 - S(T) - E(T) = q \int_0^T E(t) dt. \quad (13)$$

对模型(11)第 3 和第 4 个方程两边同时从 0 到 T 积分得

$$\int_0^T I(t) dt = \frac{pq \int_0^T E(t) dt + I_0 - I(T)}{\gamma}, \quad \int_0^T A(t) dt = \frac{(1-p)q \int_0^T E(t) dt + A_0 - A(T)}{\gamma_1}. \quad (14)$$

将(14)式代入(12)式和(13)式得到如下计算基本再生数定理.

定理 4 假设初始时刻 $I_0 = E_0 = A_0 = 0$, 染病峰值为 $I(T)$ 和累计出院数为 $R(T)$, 则 \mathcal{R}_0 是如下隐式方程的解 $N(\alpha - 1) + I(T) + A(T) + R(T) = \frac{N}{\mathcal{R}_0} \left(\ln \left(\frac{\alpha N}{S(T)} \right) + \frac{I(T)}{\gamma} + \frac{A(T)}{\gamma_1} \right)$. 这里 $S(T) = N - \frac{\gamma}{pq} I(T) - I(T) - A(T) - R(T)$, $A(T) = \frac{1-p}{p} I(T) - R(T) + \frac{\gamma}{p} \int_0^T I(t) dt$. $S(T)$, $E(T)$, $I(T)$, $A(T)$ 和 $R(T)$ 表示

在峰值 T 时刻易感者、潜伏者、染病者、无症状者和康复者的数量. $\int_0^T I(t) dt$ 表示在峰值时刻累计染病者数量.

根据文献[7], 可以选取参数 $p = 0.751$ 和 $\gamma_1 = 1/12$. 同样可以计算出在 SEIAR 模型下全国、湖北和广东新冠肺炎传播的基本再生数(见表 3).

表 3 全国、湖北及广东新冠肺炎传播基本再生数

Tab.3 Basic reproduction number of COVID-19 in China, Hubei and Guangdong

区域	$\alpha=0.2$	$\alpha=0.4$	$\alpha=0.6$	$\alpha=0.8$	$\alpha=1$
全国	$\mathcal{R}_0 = 2.399 6$	$\mathcal{R}_0 = 4.799 1$	$\mathcal{R}_0 = 7.198 7$	$\mathcal{R}_0 = 9.598 2$	$\mathcal{R}_0 = 11.99 8$
湖北	$\mathcal{R}_0 = 1.095 7$	$\mathcal{R}_0 = 2.191 3$	$\mathcal{R}_0 = 3.287 0$	$\mathcal{R}_0 = 4.382 6$	$\mathcal{R}_0 = 5.478 3$
广东	$\mathcal{R}_0 = 2.166 3$	$\mathcal{R}_0 = 4.332 6$	$\mathcal{R}_0 = 6.498 8$	$\mathcal{R}_0 = 8.665 1$	$\mathcal{R}_0 = 10.83 1$

在 SEIAR 模型框架下, 基本再生数随高易感比例呈单调上升的关系. 在同等比例下, 全国的基本再生数高于广东和湖北. 比较这三类模型, 显然在 SEIAR 模型框架下, 估计出的基本再生数明显高于 SIR 和

SEIR 模型.这主要源于 SEIAR 模型框架下增加了一条由无症状感染者感染易感人群的途径.

2 结 论

本文给出了一种完全结合数据和模型结构估计基本再生数的新方法.无论模型结构如何,基本再生数都可以利用确诊病例峰值相关数据计算出来,无须再做参数估计.这种方法即可避免参数估计带来的误差,也可避免估计参数由结构不唯一带来的问题.从本文可以看到,基本再生数的表达式无论是 SIR 模型、SEIR 模型还是 SEIAR 模型都和初始值无关,但由定理 1 至定理 4 可以看出其和易感人群的初值却有着很大的关联.随着模型结构的复杂,计算基本再生数需要的参数逐渐增多,计算结果差别较大,说明准确计算基本再生数和模型的结构有着密切的关系.另外,从表 1 至表 3 不难发现,仓室建模如何定义易感人群至关重要.

参 考 文 献

- [1] 新型冠状病毒(2019-nCov)感染肺炎防控集成信息平台[EB/OL].[2020-03-05].<http://www.clas.ac.cn/xwzx2016/163486/xx-fysjpt2020/>.
- [2] 2019 新型冠状病毒[EB/OL].[2020-03-04].[https://baike.baidu.com/item/2019 新型冠状病毒/24267858? fr=aladdin](https://baike.baidu.com/item/2019%20新型冠状病毒/24267858?fr=aladdin).
- [3] China News.Mayor of Wuhan: More than 5 million people left Wuhan[EB/OL].[2020-01-26].<http://ent.chinanews.com/sh/2020/01-26/9070484.shtml>.
- [4] LIU Z,MAGAL P,SEYDI O,WEBB G,Predicting the Cumulative Number of Cases for the COVID-19 Epidemic in China From Early Data[J].2020,<https://arxiv.org/abs/2002.12298>.
- [5] 王霞,唐三一,陈勇,等.新型冠状病毒肺炎疫情下武汉及周边地区何时复工?数据驱动的网络模型分析[EB/OL].[2020-03-04].<https://doi.org/10.1360/SSM-2020-0037>.
WANG X,TANG S Y,CHEN Y,et al.When will be the resumption of work in Wuhan and its surrounding areas during COVID-19 epidemics? A data-driven network modelling analysis[EB/OL].[2020-03-04].<https://doi.org/10.1360/SSM-2020-0037>.
- [6] SHAO P,SHAN Y G,Beware of asymptomatic transmission: study on 2019-nCov prevention and control measures based on SEIR model[EB/OL].[2020-03-05].<https://doi.org/10.1101/2020.01.28.923169>.
- [7] LUO X,FENG S,YANG J,et al.Analysis of potential risk of COVID-19 infections in China based on a pairwise epidemic model[EB/OL].[2020-03-05].<https://www.preprints.org/manuscript/202002.0398/v1>.
- [8] YANG J,XU F,The computational approach for the basic reproduction number of epidemic models on complex networks[J].IEEE Access,2019,7:26474-26479.
- [9] 中华人民共和国健康委员会[EB/OL].[2020-03-04].<http://www.nhc.gov.cn/xcs/yqtb/listgzbd.shtml>.
- [10] 中国统计年鉴 2019[EB/OL].[2020-03-04].<http://www.stats.gov.cn/tjsj/ndsj/2019/indexch.htm>.

The computational method of the basic reproduction number of COVID-19 by data-driven

Yang Junyuan¹, Zhang Shuo¹, Wang Guoqiang¹, Li Xuezhi²

(1.Complex Systems Research Center, Shanxi University, Taiyuan 030006, China;

2.College of Mathematics and Information Science, Henan Normal University, Xinxiang 453007, China)

Abstract: In 2019, emerging COVID-19 outbreak has a severe impact on the development of our society and human life. How to scientifically and reasonably compute the basic reproduction number of COVID-19 plays an important role in presentation of COVID-19 infection. Recently, the existing computational method of the basic reproduction number for COVID-19 propagation is mainly based on the parameter identification. In this paper, we propose a novel approach to directly compute the basic reproduction number by reported data of COVID-19 in China. Applying such an approach and reported cases at peak arrival time, we employ SIR model, SEIR model and SEIAR model to calculate the basic reproduction number of COVID-19 infection in China, Hubei and Guangdong. This method reflects the relationship between the basic reproduction number and reported data.

Keywords: the basic reproduction number; SIR model; SEIR model; peak value

[责任编辑 陈留院 赵晓华]



本期专家介绍



杨俊元, 山西大学复杂系统研究所教授, 博士, 硕士生导师, 山西省优秀青年学术带头人和山西省“131”领军人才. 曾先后在加拿大劳瑞尔大学做博士后研究(2012.1—2013.1)、在美国佛罗里达大学数学系做访问学者(2014.5—2015.4). 主要从事传染病动力学、复杂网络及博弈论研究. 主持国家自然科学基金 3 项、省部级项目 4 项, 先后在 *J Biol Math*, *Math Biosci*, *Nonlinear Anal: RWA*, *Chaos* 等杂志上发表 30 余篇论文. 在 Springer 出版专著 1 部.

徐久成, 河南师范大学二级教授, 博士, 博士生导师, 全国优秀教师. 现为教育部高等学校计算机类专业教学指导委员会委员(2018—2022), 河南省高校计算机教育研究会副理事长, 计算机科学与技术省一级重点学科带头人, 智慧商务与物联网技术河南省工程实验室主任, 计算智能与数据挖掘河南省高校工程技术研究中心主任. IEEE 国际粒计算委员会程序委员, 中国粗糙集与软计算专业委员会常务委员. 主要从事人工智能、粒计算与知识获取、生物信息学等领域的理论及应用研究. 主持国家自然科学基金项目 4 项、河南省高校新世纪优秀人才计划重大项目 1 项. 发表学术论文 130 余篇, 入选 ESI 高被引论文 2 篇. 出版专著 2 部, 主编国家规划统编教材 3 部. 获河南省高等教育省级教学成果一等奖 3 项, 作为主要参与者获国家教学成果二等奖 2 项.



杨宗献, 河南师范大学二级教授, 博士生导师, 河南省特聘教授, 河南师范大学物理学院院长, 国家级物理实验教学示范中心主任, 河南省物理学会副理事长. 留学美国、加拿大、瑞典博士后和客座教授. 主要研究领域为凝聚态物理、计算纳米科学与清洁能源材料设计. 主持国家自然科学基金项目 5 项、国际合作项目 1 项, 在 SCI 源期刊发表学术论文 160 篇. 获全国优秀科技工作者、河南省杰出人才、河南省政府津贴等奖励和荣誉.