基于 SEIR 模型的新冠疫情分析

李志伟田思琪崔梦韬薛璇燕明琪

2020/6/24

摘要:本文根据 2020 年 1 月 20 日到 6 月 3 日的新冠肺炎疫情数据,采用极大似然估计法,运用 SEIR 模型分别对全国、湖北、去全国除湖北以外地区和武汉地区的疫情进行实证分析。结果表明,在封城等措施实施以前,所有情形下基本再生数都大于 1,说明如果不对疫情加以控制,疫情会持续发展下去;封城等措施实施以后,所有情形下控制系数都大于 0,表明控制措施实施有效。同时表明 SEIR 模型可以很好地对疫情发展趋势进行模拟。

关键词:新冠疫情 SEIR 模型有效再生数

1. 绪论

1.1 研究背景

2019 年 12 月中旬以来,湖北省武汉市短期内出现以发热、乏力、咳嗽、呼吸不畅等为主要症状的不明原因肺炎病例。随着疫情的扩散,我国其他地区及境外也相继出现此类病例。研究发现,该疾病的元凶是一种新型冠状病毒,该病原感染所致的肺炎称为新型冠状病毒肺炎。疫情发生后,党中央和国务院高度重视,中央成立应对疫情工作领导小组。习近平总书记作出重要指示,强调要把人民群众生命安全和身体健康放在第一位,把疫情防控工作作为当前最重要的工作来抓。全国各省市也启动了突发公共卫生事件一级响应机制,采取有力措施,坚决打赢疫情防控阻击战。截至 2020 年 6 月 10 日 24 时,我国累计报告确诊病例 83057 例,现有确诊病例 62 例,国内疫情已经相对稳定。

1.2 文献综述

现有很多文章利用了模拟退火算法、极大似然估计方法以及 SEIR 模型进行研究。曹志东、王劲峰等(2009),构建了反映 SARS 传播过程的动力学模型,利用模拟退火算法对 2003 年广

州市的 SARS 疫情进行了研究,得到了潜伏期、感染期、基本再生数与有效再生数的最优估计; 王小莉、曹志东等 (2011),利用北京市 2009 年 8 月 3 日至 9 月 27 日经校正的甲型 H1N1 流感样病例数,建立 SEIR 模型,对国庆后甲型 H1N1 流感的流行趋势及高峰达到时间进行预测,与甲型 H1N1 流感病原学监测结果进行比较,评价模型的预测效果。在研究过程中,采用模拟退火算法进行流行病学参数的优化; 蔡磊 (2016),根据 2014-2015 年西非埃博拉病毒数据,采用极大似然估计方法,运用三类传染病动力学模型 (SIR 模型、SEIR 模型和 SEIHFR 模型),分别对总体、几内亚、塞拉利昂和利比里亚的疫情进行实证分析; 耿辉、徐安定等 (2020)建立 SEIR 模型,加入潜伏期传染率、感染人群变化率等新参数,分析防控手段的有效性; 黄森忠、彭志行等 (2020)基于自 2019 年 12 月 12 日以来发布的确诊数据,对新型冠状病毒肺炎疫情的流行趋势进行了研判,对疾控策略的效率进行了评估。蔡洁、贾浩源等 (2020) 收集 2020 年 1 月 20 日 2 月 18 日武汉市新冠肺炎疫情的相关数据,模拟新冠肺炎疫情发展趋势。同时,加入不同管控时效和不同管控强度等参数分析防控策略对武汉市新冠肺炎疫情发展的影响。

本文在现有研究的基础上,主要利用极大似然估计方法对 SEIR 模型中的参数进行估计,进而利用 SEIR 模型对疫情进行分析。

2 SEIR 模型与参数估计方法

2.1 传染病动力学模型基本概念

2.1.1 有效接触率和传染率

在传染病动力学模型中,传染率是一个非常重要的参数。一般来说,传染病是通过接触进行传播的。接触率被定义为单位时间内一个患者与其他成员接触的次数,它通常依赖于环境中的总人数 N。在传染病传播的环境中,一般包括易感者、患者、免疫者和潜伏者,假设患者和其他患者、免疫者和潜伏者接触不会发生传染,只与易感者接触时才会发生传染,同时假设每次接触传染的概率为 β_0 ,接触率为 C(N),易感者在总人群中的占比为 S/N,则每一患者对易感者的平均有效接触率应为 $\beta_0 C(N) S/N$,也就是每一个患者平均对易感者的传播率。于是在 t时刻单位时间内新增的感染人数为:

$$New = \beta_0 C(N) \frac{S(t)}{N(t)} I(t)$$

疾病发生率一般存在两种形式,一种为双线性发生率,即接触率与总人数呈正比时的发生率,此时 $New=\beta S(t)I(t)$,其中 $\beta=\beta_0\times k$;另一种为标准型发生率,即接触率为一个常数的发生率,此时 $New=\beta S(t)/N(t)I(t)$, $\beta=\beta_0\times k$ 。有专家支出,对于人类和某些群居动物而言,标准发生率比双线性发生率更符合实际。在经典的传染病模型中,这两种传染率都被大量使用。

2.1.2 再生数

对于经典的传染病动力学模型来说,再生数 R 是非常重要的参数。再生数表示当总人数达到平稳企鹅个体均为易感者时,引入一个患者,在平均感染期内所能传染的人数。因此,R=1可以作为衡量疾病是否得到控制的阈值,当 R<1 时,即一个感染者在平均感染期内传染的最大人数小于 1 时,此时就算不对疾病施加任何控制措施,疫情也会逐渐消失;当 R>1 时,疫情将会逐渐传播开来从而造成严重后果,此时只有施加控制措施使 R 降到 1 以下时,疫情才会得到有效控制,从而逐渐消失,因此,对再生数的分析时传染病模型的重要工作。

根据相关文献,再生数可以用传染率和平均感染期的乘积表示,即 $R = \frac{\beta}{\gamma}$,其中 β 表示传染率, $1/\gamma$ 表示平均感染期。

2.2 SEIR 模型和参数设计

2.2.1 SEIR 模型

SEIR 模型是 Aron 在 1984 年首次提出,与 SIR 模型相比,在易感者和感染者之间增加了潜伏者这一类,也就是说 SEIR 模型将人群分为四个状态: 易感者 (S)、潜伏者 (E)、感染者 (I)、移出者 (R)。易感者是指没有患病,但缺乏免疫力,与感染者接触后易患病的人;潜伏者是指已经感染,但未表现出症状的人;感染者是指已经受病毒感染,并且有明显的症状表现,可以传染给易感者的人; 移出者是指死亡或者病愈的人群。

该模型中做出了如下假设:

- (1) 假设所研究地区所研究的人口总数为 N,并且不随时间的变化而变化,不考虑自然出生率、死亡率等因素;
- (2)假设易感者由于受传染病的影响,其人数随时间变化的变化率与当时易感者人数和感染者人数成正比:
 - (3) 假设感染者死亡或治愈转移到移出者人群的转换率与当时感染者人数成正比。

SEIR 模型易感者人群从患病到移出的具体传染过程如下所示:

$$S \to E \to I \to R$$

其中:

过程 $S \to E$ 表示易感者人群接触到病毒的感染转变为潜伏者人群;

过程 $E \rightarrow I$ 表示易感者人群经历平均潜伏期后变为病毒感染者人群;

过程 $I \to R$ 表示感染者人群经过平均感染期 $1/\gamma$ 后死亡或者治愈的移出者人群。

2.2.2 参数设计

1. 传染率设计。

由于传染并发展到一定的时期,将会受到控制措施的影响,这种控制措施只会影响感染者的传播过程,抑制其传播。在控制措施实施之前,传染率 $\beta(t)$ 是一个常数,即 $\beta(t)=\beta_0$;在实施之后,因为这种控制措施的实施效果不是瞬间实现的,本文根据相关理论将感染率设置为呈指数下降,即在 $t \geq \tau$ 时:

$$\beta(t) = \beta_0 e^{[-k(t-\tau)]}$$

其中 τ 表示控制措施实施的时间,k 表示在 $t \ge \tau$ 时,传染率 $\beta(t)$ 呈指数下降的系数。

2. 再生数设计。

再生数时表示在完全易感人群中一个患者在整个传染期平均感染的人数。它依赖于传染率和平均感染期,而感染率是受政府控制措施的影响,并且在政府控制措施实施前传染率是不同的。在实施之前,传染率是一个不变的常数 β_0 ; 在实施之后,传染率将呈指数下降,此时传染率为 $\beta_0e^{[-k(t-\tau)]}$ 。因此,本文将再生数分为基本再生数和有效再生数两类。

- (1) 基本再生数。基本再生数是指在控制措施实施之前,完全易感人群中一个患者在整个传染期内平均感染的人数,此时再生数为一个恒定的常数,反映的是疫情发生初期的发展趋势,此时 $R_0=\beta_0/\gamma$,其中 β_0 表示控制措施实施之前的传染率, $1/\gamma$ 表示感染者从感染开始到移出时的平均感染时间。若基本再生数大于 1,表明疫情会继续发展。
- (2) 有效再生数。有效再生数是指在控制措施实施之后,一个患者在整个传染期内平均感染的人数,即 $R(t) = \beta(t)/\gamma$,其中 $\beta(t)$ 表示控制措施实施之后的传染率, $1/\gamma$ 表示感染者从感染开始到移出时的平均感染时间。在一个封闭的人口规模中,由于易感者人群在不断减少,所以有效再生数是非增的,当有效再生数降到 1 以下是,表明控制措施的实施是有效的,即控制措施对疫情的发展产生了有效的抑制作用。

2.3 极大似然估计法

极大似然法,也称最大似然法,最早由德国数学家高斯于 1821 年提出,在 1912,英国统计学家罗纳德费歇尔再次提出了此方法的思想,并证明了该方法的一些性质,自此得到广泛应用和研究。

极大似然法是从极大似然原理出发发展起来的其他参数估计方法的基础。对于极大似然法, 当从模型总体中随机抽取 n 组样本观测值后,最合理的参数估计量应该使得从此模型中抽取该 n 组样本观测值的概率最大。 设样本分布的密度函数为 $f(x,\theta)$,对给定的 x,可把 $f(x,\theta)$ 看成是 θ 的函数,则 $f(x,\theta)$ 为定义在参数空间上的似然函数:

$$L(\theta, x) = f(\theta, x), \theta \in \Theta, x \in X \subseteq \mathbb{R}^n$$

在求取 MLE 时,如果似然函数是连续可导的,则参数 θ 的估计值 $\hat{\theta}$ 可由以下方程求出,否则应按定义来取:

$$\frac{\alpha L(\theta, x)}{\alpha \theta} = 0$$

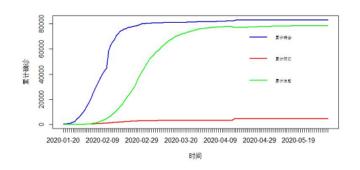
3 数据来源及描述性统计

3.1 数据来源

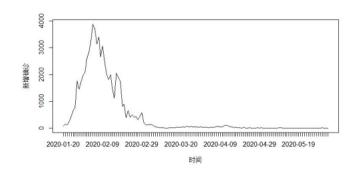
此次研究所使用的数据来源于狗熊会团队手动整理的数据。数据的指标包括:新增确诊人数、累计确诊人数、现有重症人数、累计死亡人数、累计治愈人数、累计密切解除人数、正在接受医学观察人数、新增密切解除人数、累计报告确诊病例省(区、市)数量。每条数据表示从2020年1月20日至2020年6月3日期间每日疫情的发展情况。

3.2 描述性统计分析

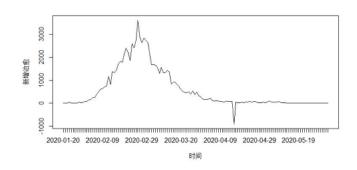
自 2020 年 1 月 20 日至 2020 年 6 月 3 日为止,我国新冠肺炎的累计确诊人数 83022 人. 如图所示,将累计确诊人数、累计治愈人数和累计死亡人数放在同一个图中。从图中可以看出。新冠肺炎的累计确诊人数在 2020 年 1 月 20 日至 2020 年 1 月 27 日缓慢增长,在 2020 年 1 月 27 日之后我国的新冠肺炎的累计确诊人数以斜率先增加后减小的形式增加,在 2020 年 2 月 27 日之后我国新冠肺炎的累计确诊人数趋于平稳。累计治愈人数先以斜率增大的形式增加,后以斜率减小的形式增加,在 2020 年 4 月 4 日之后趋于平缓。累计死亡人数先以斜率增大的形式增加,后以斜率减小的形式增加,在 2020 年 4 月 10 日趋于平稳。但在 2020 年 4 月 16 日发生了跳跃,这是因为在 2020 年 4 月 17 日对 2020 年 4 月 16 日的湖北省武汉市新冠肺炎疫情数据进行了订正,累计死亡人数核增 1290 例。累计治愈人数的曲线远远高于累计死亡人数的曲线,累计治愈人数和累计确诊人数的曲线是非常接近的,由此可知,我国新冠肺炎的治愈率还是比较高的,达 94%。



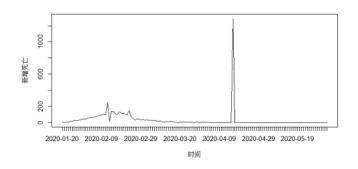
由下图可知,我国新冠肺炎的新增确诊人数先快速增加,在 2020 年 2 月 4 日,达到最高点,新增确诊人数为 3887 例。在 2020 年 2 月 4 日之后,我国新冠肺炎的新增确诊人数开始下降,直至 2020 年 3 月 1 日,我国新冠肺炎的新增确诊人数趋于平缓。



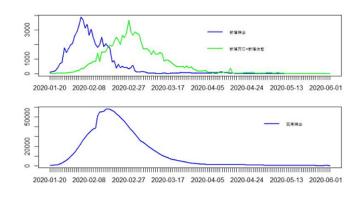
由下图可知,我国新冠肺炎的新增治愈人数先呈现上升的趋势,在 2020 年 2 月 26 日,达到最高点,新增治愈人数为 3622 例。在 2020 年 2 月 26 日之后,我国新冠肺炎的新增治愈人数开始下降,直至 2020 年 3 月 30 日,我国新冠肺炎的新增治愈人数趋于平缓。在 2020 年 4 月 16 日,我国我国新冠肺炎的新增治愈人数出现了负值,这是由于在 2020 年 4 月 17 日对 2020 年 4 月 16 日的湖北省武汉市新冠肺炎疫情数据进行了订正,而新增治愈人数是通过使用原始数据的累计治愈人数的一阶差分得到的,这就使得 2020 年 4 月 16 日的新增治愈人数受到了疫情数据订正的影响,2020 年 4 月 16 日的实际新增治愈人数为 52 例。



如图所示,我国新冠肺炎的新增死亡人数先呈现上升的趋势,在 2020 年 2 月 12 日,达到最高点,新增死亡人数为 254 例。在 2020 年 2 月 12 日之后,我国新冠肺炎的新增死亡人数开始下降,直至 2020 年 2 月 24 日,我国新冠肺炎的新增死亡人数趋于平缓。在 2020 年 4 月 16 日,我国我国新冠肺炎的新增死亡人数出现了异常的趋势,这是由于在 2020 年 4 月 17 日对 2020 年 4 月 16 日的湖北省武汉市新冠肺炎疫情数据进行了订正,而新增死亡人数是通过使用原始数据的累计死亡人数的一阶差分得到的,这就导致 2020 年 4 月 16 日的新增死亡人数受到了疫情数据订正的影响,2020 年 4 月 16 日的实际新增死亡人数为 0 例。



将新增确诊人数、新增治愈人数 + 新增死亡人数放形成的曲线图和现有确诊的曲线图放在一起形成图 8. 由图可知,自 2020 年 1 月 20 日至 2020 年 2 月 18 日,新冠肺炎的新增人数确诊大于新增治愈和新增死亡的总人数,此时现有确诊人数上升。自 2020 年 2 月 18 日起,新增确诊人数小于新增治愈和新增死亡的总人数,此时现有确诊人数下降。



4 基于 SEIR 模型疫情实证分析

根据全国不同地区疫情发展的不同趋势,本文将分别对全国、湖北、全国除湖北以外地区以及武汉疫情发展趋势进行分析。对于 SEIR 模型,首先根据我国不同地区新冠肺炎疫情发展态势,做出一些基本假设:

- (1)假设全国总人口规模 $N_0=1.4\times 10^9$,湖北总人口规模 $N_1=5\times 10^7$,全国除湖北以外地区总人口规模 $N_2=1.35\times 10^9$,武汉人口总规模 $N_3=1.1\times 10^7$;初始各个区域只有一个感染者,没有潜伏者,也没有移出者,即状态空间用 C_0 表示, $C_0=(S_0,E_0,I_0,R_0)=(N_i-1,0,1,0)$;
- (2) 根据相关研究,设定平均潜伏期 $\frac{1}{\delta}$ 和平均传染期 $\frac{1}{\gamma}$,并假设传染率 β 满足 $0 < \beta < 1$,死亡率 f 满足 0 < f < 1 参数设定如下表所示:

地区	全国	湖北	全国除湖北	武汉
$\frac{1}{\delta}$	4.98	4.77	6.18	4.45
$\frac{1}{\gamma}$	7.04	6.74	6.24	4.86

- (3) 假设控制措施是在 1 月 23 日武汉开始封城时开始实施的;
- (4) 假设患者的累计人数服从泊松分布。

根据新冠肺炎的传染机制,对新冠疫情建立微分方程模型如下:

$$\begin{cases} \frac{dS}{dt} = & -\beta e^{(-kt)}S(t)I(t)/N \\ \frac{dE}{dt} = & \beta e^{(-kt)}S(t)I(t)/N - \sigma E(t) \\ \frac{dI}{dt} = & \sigma E(t) - \gamma I(t) \\ \frac{dR}{dt} = & (1-f)\gamma I(t) \\ \frac{dD}{dt} = & f\gamma I(t) \\ \frac{dC}{dt} = & \sigma E(t) \end{cases}$$

其中,N 表示总人口规模, β 表示感染者人群的传染率,I/N 表示与感染者接触的概率; $\frac{1}{\delta}$ 表示平均潜伏期, $\frac{1}{\gamma}$ 表示平均感染期,f 表示感染者人群的死亡率。在 t 时刻的数量分别记为 S(t), E(t), I(t), R(t),且 S(t) + E(t) + I(t) + R(t) = N。C(t) 表示从感染者病症开始到 t 时刻所有病例的累计人数,D(t) 表示在 t 时刻所有感染者死亡的人数,并且假设 C(t) 和 D(t) 均服从于泊松分布。

在给定初始状态以及初始参数的条件下,利用 SEIR 模型,结合极大似然估计方法,以及 Nelder-Mead 最优算法,对待估参数进行估计,最后利用估计的参数值进行模拟并与原始数据 拟合比较。

对于极大似然估计方法,这里是通过病例的累计人数 C 服从泊松分布,并利用累计人数的模拟数据作为参数来建立似然函数的。

此时似然函数为:

$$L = \Sigma \frac{\hat{C}^C}{C!} e^{-\hat{C}}$$

其中 \hat{C} 代表患者累计人数的模拟序列,C 代表患者累计人数的原始序列。

4.1 重要参数分析

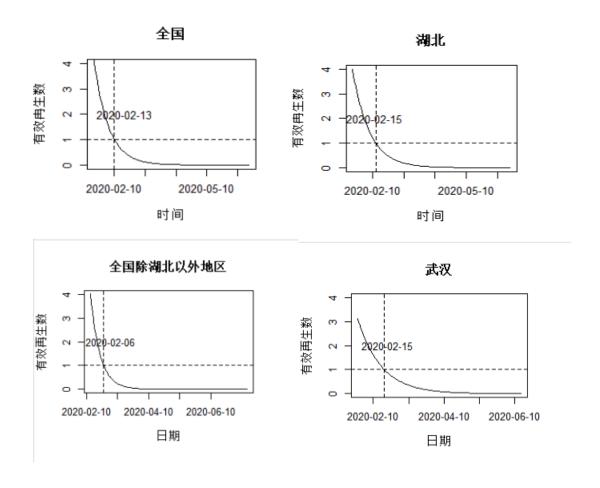
根据 SEIR 微分方程模型,结合上述假设,采用极大似然估计法,并结合 Nelder-Mead 最优算法,得出疫情参数估计结果:

参数	β	k	δ	γ
全国	0.617	0.072	0.201	0.142
湖北	0.591	0.062	0.210	0.148
全国除湖北	0.644	0.108	0.162	0.160
武汉	0.644	0.051	0.225	0.206

由上表可以看出,全国、湖北、全国除湖北和武汉地区的新冠疫情传染率分别为 0.617、 0.591/0.644 和 0.644; 基本再生数 R_0 表示在干预之前,完全易感人群中一个病例在整个传染期内平均感染的人数,即 $R_0 = \beta/\gamma$,四种情形下基本再生数 R_0 分别为 4.344、3.982、4.17 和 3.128,可以看出,四种情形的基本再生数均大于 1,这说明在控制措施干预之前,新冠疫情会持续蔓延,此时只有政府实施控制措施干预使其降低至小于 1 时,疫情才会得到有效抑制,并最终慢慢消失。

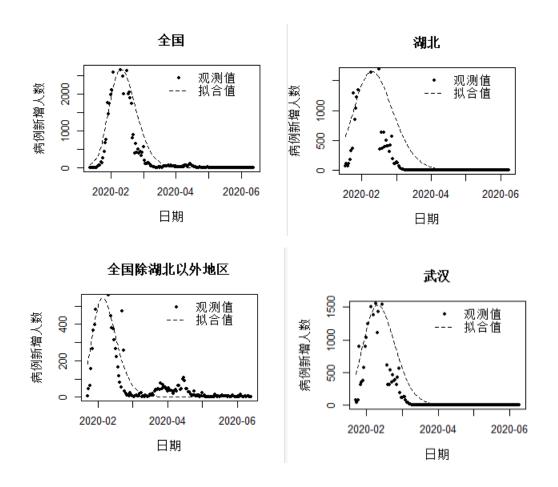
4.2 有效再生数分析

对于控制系数 k 来说,四种情形下,疫情的控制系数分别为 0.072、0.062、0.108 和 0.051,都是大于 0 的。可以看出通过政府的控制干预措施能够有效的抑制新冠疫情,可以通过有效再数指标来进一步体现措施干预后疫情的抑制情况。有效再生数表示在控制措施干预以后,一个病例在整个传染期内平均感染的人数,即 R(t) 为 $\beta e^{(-k(t-\tau))}$ 。由下图可以看出,有效再生数都是呈指数下降趋势,对于全国疫情,有效再生数下降到 1 的日期为 2020 年 2 月 13 日,对于湖北疫情,有效再生数下降到 1 的日期为 2020 年 2 月 15 日,对于全国除湖北以外地区疫情,有效再生数下降到 1 的日期为 2020 年 2 月 15 日,对于武汉疫情,有效再生数下降到 1 的日期为 10000 年 1000 年 100 年 100 日 10



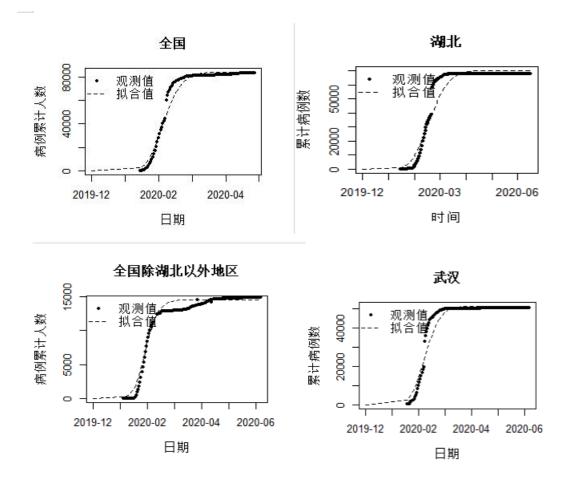
4.3 新增病例人数拟合

通过对重要参数的估计和分析,利用相关参数,结合 SEIR 模型模拟出新增病例人数,并与真实新增病例人数进行对比拟合,如下图所示,我们发现四种情形下疫情的新增病例人数观测值与拟合值的发展趋势基本一致,并且拟合曲线基本吻合。对于全国疫情,可以看出新增病例拟合值在 2020 年 2 月 3 日达到峰值,随后开始逐日下降,而实际新增确诊人数也是从 2 月 4 日左右开始有下降趋势的;对于湖北疫情,新增病例数的拟合值在 2020 年 2 月 8 日达到峰值,随后开始逐日下降,而真实值也是从 2 月 10 日左右开始有下降趋势的;对于全国除湖北以外地区疫情,理论新增确诊人数在 2020 年 2 月 5 日达到峰值,随后开始逐日下降,而实际的当时新增确诊人数是从 2 月 5 日开始有下降趋势的;对于武汉疫情,理论当日新增确诊人数在 2020 年 2 月 23 日达到峰值,随后开始逐日下降,而实际的当时新增确诊人数是从 2 月 20 日左右开始有下降趋势的。也就是说,四种情形下的理论观点分别出现在 2 月 3 日、2 月 8 日、2 月 5 日和 2 月 23 日。



4.4 累计病例数拟合分析

通过上述重要参数的估计及分析,利用参数的估计值,结合 SEIR 微分方程模型模拟出新冠肺炎病例的累计人数的拟合值,并做出相应曲线与真实观测值进行对比拟合。如图所示。我们发现四种情形下疫情的真实观测值与拟合值的发展趋势基本一致,并且拟合曲线基本吻合,这表明 SEIR 模型对全国、湖北、全国除湖北以外地区以及武汉的新冠肺炎疫情的传播机制和传播速度能够给出有效的解释和说明。



5 结论

综上所述,本文收集整理了 2020 年 1 月 20 日至 6 月 3 日的确诊、新增、死亡、治愈病例 人数等官方网络疫情通报数据,采用极大似然估计法,并结合 Nelder-Mead 最优算法对 SEIR 传染病动力学模型的相关参数进行估计,分别得出全国、湖北省、除湖北省以外全国其余省份以及武汉市的疫情模型参数估计结果。通过建立 SEIR 模型对这四个板块疫情发展规律进行了分析,估计了防控隔离和政府干预等措施对疫情发展的影响。分析结果表明:政府的干预措施是有效遏制疫情大面积传播的重要举措。政府倡导下的居家防控隔离措施对疫情大面积传播有重要抑制作用,严格对感染者所接触过的人员进行医学追踪隔离等措施也有效地防止了疫情的过快增长。

本文基于极大似然估计法建立的 SEIR 模型对新冠疫情进行回溯分析,研究结果确认了该模型在传染病传播态势分析上是可靠的。但是,模型不可避免地会与现实存在一定差异,进而导致分析结果存在一定的偏差,希望大家批评指正。

参考文献

- [1] 曹志冬, 王劲峰, 韩卫国, 高一鸽, 曾光. 广州市 SARS 流行的数学建模与干预措施的定量评估 [J]. 复旦学报 (自然科学版),2009,48(06):793-800.
- [2] 王小莉, 曹志冬, 曾大军, 杨鹏, 张奕, 钱海坤, 田丽丽, 黎新宇, 王全意. 应用 SEIR 模型预测 2009 年甲型 H1N1 流感流行趋势 [J]. 国际病毒学杂志,2011(06):161-165.
- [3] 蔡磊. 基于 2014-2015 年埃博拉疫情数据的统计建模与重要参数分析 [D]. 暨南大学,2016.
- [4] 耿辉,徐安定,王晓艳,张勇,尹小妹,马茂,吕军.基于 SEIR 模型分析相关干预措施在新型 冠状病毒肺炎疫情中的作用 [J]. 暨南大学学报 (自然科学与医学版),2020,41(02):175-180.
- [5] 黄森忠, 彭志行, 靳祯. 新型冠状病毒肺炎疫情控制策略研究: 效率评估及建议 [J]. 中国科学: 数学,2020,50(06):885-898.
- [6] 蔡洁, 贾浩源, 王珂. 基于 SEIR 模型对武汉市新型冠状病毒肺炎疫情发展趋势预测 [J]. 山东 医药,2020,60(06):1-4.