

基于变系数 Hawkes 过程的新型 冠状病毒肺炎的传播强度分析*

吴远山[†]

(中南财经政法大学统计与数学学院, 武汉 430073)

(E-mail: wu@zuel.edu.cn)

黄佳雨

(武汉大学数学与统计学院, 武汉 430072)

(E-mail: jerry_huang@whu.edu.cn)

叶桢均

(中南财经政法大学统计与数学学院, 武汉 430073)

(E-mail: bigyeanjun@163.com)

摘要 当前, 新型冠状病毒肺炎 (COVID-19) 在全球大部分地区肆虐, 对人类生命安全和
社会经济活动带来严峻挑战. 新型冠状病毒肺炎的传播强度随时间的变化规律及其影响因素是传
染病学家关心的重要问题. 基于新型冠状病毒肺炎病例数, 本文提出变系数 Hawkes 过程, 分
析日本, 韩国以及北京市, 武汉市新冠肺炎传播强度随时间变化的趋势, 评估突发事件和政府
采取的干预措施对疫情防控的影响. 本文基于样条逼近技术建立半参数估计方法, 数值模拟结
果表明所提出方法在有限样本下具有良好表现. 该研究可为疫情防控提供有益的统计建议.

关键词 点过程; 复发事件; 半参数方法; 样条函数; 拐点

MR(2000) 主题分类 60G55; 60G35

中图分类号 O211.65

1 引言

2019 年末, 新型冠状病毒肺炎 (COVID-19, 以下简称新冠肺炎) 在湖北省武汉市华

本文 2020 年 12 月 22 日收到, 2021 年 4 月 2 日收到修改稿.

* 国家自然科学基金 (12071483) 资助项目.

[†] 通讯作者.

南海鲜市场被发现并相继在我国及世界各地蔓延。经政府全面积极地采取预防控制措施和医疗救援措施,我国疫情得到有效的控制,目前全国范围内各项社会经济活动和人们日常生活恢复正常。然而,中国以外的部分国家和地区的疫情发展形势依然严峻。截至2021年4月14日,全球新冠肺炎累计确诊人数突破一亿三千万,死亡人数二百九十余万,给人类生命安全和社会经济活动带来严重挑战(<https://covid19.who.int/table>)。

自疫情蔓延以来,国内外众多学者对其发展态势进行分析和预测。Gao等^[1]采用基于玻尔兹曼函数的模型预测中国新冠肺炎累计死亡人数;Wu等^[2]采用SEIR模型分析疫情发展,并运用马尔科夫链蒙特卡罗方法进行模拟预测;Zhou等^[3]进行回顾性队列研究,建立单变量和多变量的logistic回归模型探讨与住院患者死亡率相关的影响因素。Yu等^[4]结合SIR模型与SEIR模型对疫情进行宏观分析,并利用机器学习方法对疫情规模进行预测;Wang等^[5]采用平滑指数模型对湖北省新冠肺炎病例数进行拟合和预测,证实防疫措施的有效性;Wang等^[6]建立武汉市及周边疫情严重城市的新冠肺炎传播的复杂网络模型,讨论武汉市及周边地区复工的可能时间节点和复工对二次暴发风险的影响。

为何新冠肺炎的新增病例数在一时间段内呈上升趋势,而在另一时间段内则显下降趋势?人们期盼的拐点何时到来?这些问题与新冠肺炎的传播强度随时间变化的规律及其影响因素密切相关,是传染病学家关心的重要问题。作为一种典型的当前病例(事件)发生依赖于历史病例的点过程,自激发性点过程^[7]因其聚集簇的解释^[8]能够刻画病例之间的传播强度变化规律。Kelly等^[9]利用经典的自激发性点过程拟合与预测刚果民主共和国的埃博拉病毒疫情。Huang和Wu^[10]利用Fusion Lasso的方法以检测分段线性的经典的自激发性点过程的变点并将之应用于刚果民主共和国的埃博拉病毒疫情变化的研究。然而,疾病的传播强度通常是随时间的推移而发生变化,尤其是管理者如果能采取有效的防疫措施,控制疾病的传播途径,疾病的传播强度也会相应降低。因此,考虑病例之间的传播强度随时间的变化规律具有重要的现实意义。本文提出变系数Hawkes过程以刻画新冠肺炎的传播强度随时间的变化规律。基于样条逼近的技术,本文建立了半参数的估计方法。进一步,数值模拟的结果表明所提出的估计方法在有限样本下具有良好的表现。本文分别分析了日本,韩国以及北京市,武汉市新冠肺炎的病例数据,揭示了新冠肺炎的传播强度随时间的变化规律,并研判了政府采取的防疫干预措施以及一些突发事件对该变化规律的影响。

本文的章节安排如下。在第2节中,介绍经典的Hawkes过程以及提出变系数Hawkes过程,并建立参数和非参数的估计方法。在第3节中,对本文提出的变系数Hawkes过程的模型参数的估计方法进行模拟实验。在第4节中,运用所提出的方法分析日本,韩国与北京市,武汉市的新新冠肺炎病例数据。第5节对全文进行总结与讨论。

2 估计方法

本节首先介绍经典Hawkes过程及其参数的解释,再引申出本文提出的变系数Hawkes过程,最后对变系数Hawkes过程的模型参数的估计方法进行阐述。

2.1 经典 Hawkes 过程

设 $(\Omega, \mathcal{F}_t, P)$ 是一个完备概率空间, $\mathbb{F} = \{\mathcal{F}_t\}_{t \geq 0}$ 是此空间上的一个右连续 σ -代数流. 记 $\{T_1, \dots, T_i, \dots\}$ 为一系列关于 \mathbb{F} 适应的复发事件发生时间的序列, 其中 T_i 是第 i 个事件发生的时间. 因此, $0 < T_1 < \dots < T_i < \dots$ 是一列 \mathcal{F}_t -停时. 定义计数过程 $N(t)$ 为在时间点 t 及之前所发生的累计事件数, 即

$$N(t) = \sum_{i \geq 1} I(T_i \leq t),$$

其中 $I(\cdot)$ 为示性函数. 定义截至时间点 t 之前的历史事件集合为 $\mathcal{H}_t = \{T_1, \dots, T_n \mid T_n < t\} = \{T_1, \dots, T_{N(t-)}\}$, 则给定历史事件集合时的条件强度函数可表示为

$$\lambda(t \mid \mathcal{H}_t) = E\left(\frac{dN(t)}{dt} \mid \mathcal{H}_t\right). \quad (2.1)$$

若此条件强度函数具有如下形式, $\lambda(t \mid \mathcal{H}_t) = \mu + \beta \int_0^{t-} g(t-v) dN(v)$, 其中 $\mu \geq 0$, $\beta \geq 0$ 为参数, $g(\cdot)$ 为非负的引发函数, 则称点过程 $N(t)$ 是 Hawkes 点过程. 它是一类自激发性点过程. 直观地, 条件强度函数的表达式可表明历史事件的发生对未来事件的发生具有促进作用. 这是由于此点过程的每一次跳跃会使条件强度函数增大, 进而使得再一次发生跳跃的概率增加, 从而解释了其自激发性.

2.2 变系数 Hawkes 过程

根据经典 Hawkes 过程的聚集簇解释, (2.1) 中的参数 μ 表示基准强度, 它刻画由外来因素导致事件发生的强度. 例如, 在新冠肺炎疫情中, 由于接触携带病毒的进口冷链食品外包装而产生的确诊病例和外来输入病例. 引发函数 $g(\cdot)$ 描述的是在已有单簇中下一个确诊病例产生时间的分布, 通常随时间的推移而递减. 参数 β 表示传播强度, 衡量了系统中历史事件列对当前事件发生强度的影响.

在实际中, 新冠肺炎的传播强度通常随时间的变化而变化. 在没有采取疫情防控措施的情况下, 新冠肺炎的传播强度会随确诊病例的增加而增强; 若采取紧急防控措施, 例如隔离疑似与确诊患者, 限制城市交通与社区出行等, 可对新冠肺炎传播强度产生抑制作用. 因此, 经典 Hawkes 过程中 β 与时间无关的情形不足以刻画新冠肺炎的传播强度随时间的变化而变化的规律. 鉴于此, 本文提出变系数 Hawkes 过程,

$$\lambda(t \mid \mathcal{H}_t) = \mu + \beta(t) \int_0^{t-} g(t-v) dN(v), \quad (2.2)$$

其中 μ 为待估的非负基准强度参数, 传播强度 $\beta(\cdot)$ 为待估的非负函数型参数, 它衡量了系统中历史事件列对当前事件发生强度的影响及其随时间的变化方式. 为方便计, 本文假定 $g(t) = \exp(-\alpha t)I(t > 0)$, 其中参数 α 为待估的未知参数.

2.3 参数估计

记数据集为 $\{T_1, \dots, T_n\}$. 定义 $f(s | \mathcal{H}_t)$ 与 $F(s | \mathcal{H}_t)$ 分别为给定历史事件集合 $\mathcal{H}_t = \{T_1, \dots, T_{N(t-)}\}$ 时, 下一个确诊病例产生时间 $T_{N(t)}$ 的条件密度函数与条件分布函数. 通过基本的代数运算, 可得如下等式,

$$f(s | \mathcal{H}_t) = \lambda(s | \mathcal{H}_t) \exp \left\{ - \int_{T_{N(t-)}}^s \lambda(v | \mathcal{H}_t) dv \right\}, \quad s \geq t,$$

与

$$F(s | \mathcal{H}_t) = 1 - \exp \left\{ - \int_{T_{N(t-)}}^s \lambda(v | \mathcal{H}_t) dv \right\}, \quad s \geq t. \quad (2.3)$$

基于 $\{T_1, \dots, T_n\}$ 的似然函数为

$$L(\mu, \beta(\cdot), \alpha) = \prod_{i=1}^n f(T_i | \mathcal{H}_{T_i}) = \exp \left\{ - \int_0^{T_n} \lambda(s | \mathcal{H}_s) ds \right\} \prod_{i=1}^n \lambda(T_i | \mathcal{H}_{T_i}).$$

为表达形式上的简便, 记 $Z_\alpha(t) = \int_0^{t-} g(t-v) dN(v)$. 从而, 在变系数 Hawkes 过程模型 (2.2) 的假设下, 可推导负对数似然函数为

$$l(\mu, \beta(\cdot), \alpha) = - \sum_{i=1}^n \log \{ \mu + \beta(T_i) Z_\alpha(T_i) \} + \sum_{i=1}^n \int_{T_{i-1}}^{T_i-} \{ \mu + \beta(t) Z_\alpha(t) \} dt, \quad (2.4)$$

其中 $T_0 = 0$. 为了估计函数型参数 $\beta(\cdot)$, 本文运用样条技术对其逼近. 在观测的时间区间上取一结点序列 $\{x_1, \dots, x_K\}$, 并以此形成的 r 阶样条基函数, 记为 $\{B_k(\cdot)\}_{k=1}^{K-r}$. 这里 K 和 r 为正整数, 且 $K > r$. 将 $\beta(\cdot)$ 用样条基函数的线性组合 $\sum_{k=1}^{K-r} \theta_k B_k(\cdot)$ 进行逼近, 其中 $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{K-r})^T \geq \mathbf{0}$ 为待估参数. 这里 $\mathbf{0}$ 是零向量, $\boldsymbol{\theta} \geq \mathbf{0}$ 即指 $\boldsymbol{\theta}$ 的每一分量均大于或等于 0. 该约束可保证样条估计的非负性. 将此线性组合代替 $\beta(\cdot)$, 并代入 (2.4), 仍记之为 $l(\mu, \boldsymbol{\theta}, \alpha)$. 由于 $l(\mu, \boldsymbol{\theta}, \alpha)$ 关于参数 $(\mu, \boldsymbol{\theta}, \alpha)$ 为凸函数, 则其估计可由 $(\hat{\mu}, \hat{\boldsymbol{\theta}}, \hat{\alpha}) = \arg \min_{\mu \geq 0, \boldsymbol{\theta} \geq \mathbf{0}, \alpha \geq 0} l(\mu, \boldsymbol{\theta}, \alpha)$ 给出.

3 模拟实验

本节对变系数 Hawkes 过程的估计方法进行模拟实验, 以考察其在有限样本下的表现. 取观测时间区间 $[0, 300]$, 根据变系数 Hawkes 过程 $N(t)$ 的条件强度函数 (2.2), 逐次生成 $0 < T_1 < \dots < T_n$, 其中 $n = N(300)$. 生成数据集的方式采用逆变换法. 由 (2.3), 给定历史事件集合时下一次事件发生时间的条件分布函数为

$$F(s | \mathcal{H}_t) = 1 - \exp \left\{ - \mu(s - T_{N(t-)}) - \int_{T_{N(t-)}}^s \beta(v) \int_0^{v-} g(v-u) dN(u) dv \right\}, \quad s \geq t.$$

独立生成区间 $[0, 1]$ 上均匀分布的随机变量 $\{U_i\}_{i=1}^\infty$, 下一次事件发生时间 $T_{N(t-)+1}$ 由方程 $F(s | \mathcal{H}_t) = U_{N(t-)+1}$ 求解 s 即得. 设定 $\mu = 3$ 和 $\alpha = 1$, 并考虑如下四种 $\beta(\cdot)$ 的函数形式.

(a) 二次函数

$$\beta(t) = 0.6 - 0.6 \left(\frac{t-150}{150} \right)^2;$$

(b) 指数单峰函数

$$\beta(t) = \frac{3}{2\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{t-150}{50} \right)^2 \right\};$$

(c) 平方根函数

$$\beta(t) = \frac{1}{2} \sqrt{\frac{t}{300}};$$

(d) 正弦函数

$$\beta(t) = 0.3 + 0.3 \sin \left(\frac{t}{40} \right).$$

在模拟实验中样条逼近时所选取的 $K = K_0 + 2r$ 个结点为另一组独立生成的模拟样本 $\{\tilde{T}_1, \dots, \tilde{T}_n\}$ 的 $k/(K_0+1)$ 样本分位点, $k = 0, \dots, K_0+1$, 其中 K_0 个内部结点不重复, 2 个外部结点重复 r 次. 在模拟实验中统一固定内部结点数为 $K_0 = 4$ 个, 样条阶数为 $r = 4$ 阶. 模拟重复 200 次, 关于参数 μ 与 α 的估计结果见表 1. 可见该估计量是渐近无偏的.

表 1 四种情形下 μ 与 α 的估计的偏差与标准差

情形	$\hat{\mu}$		$\hat{\alpha}$	
	偏差	标准差	偏差	标准差
(a)	0.0264	0.2774	-0.0092	0.0425
(b)	-0.1117	0.1829	-0.0431	0.0951
(c)	0.1149	0.2616	-0.0173	0.0747
(d)	-0.0813	0.2057	-0.0231	0.0714

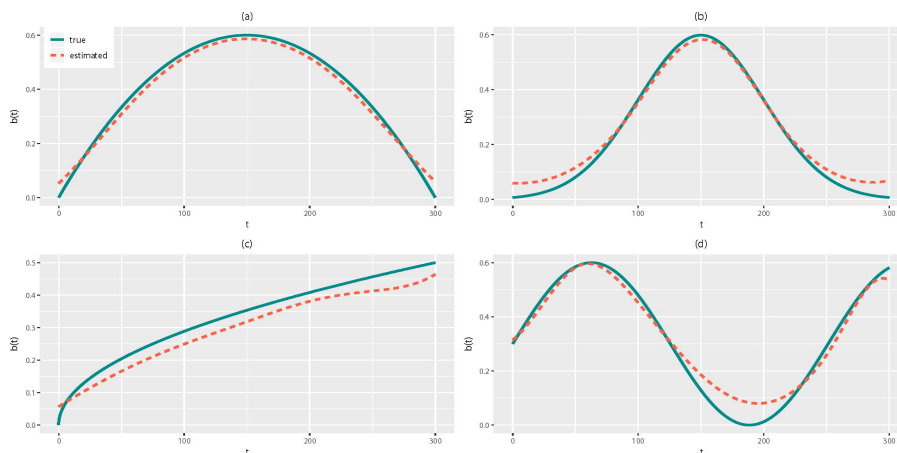
图 1 四种情形下 $\beta(\cdot)$ 的样条估计. 实线表示真实的函数曲线, 虚线表示样条估计的函数曲线

图 1 展示了 $\beta(\cdot)$ 的样条估计. 虽然呈现出一定的偏差, 但是其函数曲线的趋势与真实曲线基本吻合. 综上所述, 所提出的估计方法在有限样本下表现良好.

4 疫情数据集分析

日本, 韩国和北京市的新冠肺炎的每日确诊病例数来源于丁香医生在其官方网站所公布的数据 (<https://ncov.dxy.cn/ncovh5/view/pneumonia?from=singlemessage&isappinstalled=0>). 由此收集日本 (2020.1.15–2020.5.19), 韩国 (2020.1.2–2020.5.19) 与北京市 (2020.6.2–2020.7.2) 每日新增确诊病例数据集以及每日累计确诊病例数据集. 从湖北省卫生健康委员会于其官方网站 (<http://wjw.hubei.gov.cn>) 每日所公布的数据上收集武汉市 (2019.12.8–2020.5.13) 的每日新增确诊病例数据集以及每日累计确诊病例数据集. 本节将变系数 Hawkes 过程应用于上述数据集的分析.

日本首例患者确诊于 1 月 15 日, 该患者于 1 月 6 日前往神奈川县医疗机构就诊, 日本国立感染症研究所 15 日晚经检测确认, 患者样本新型冠状病毒呈阳性. 截至 5 月 19 日, 日本新冠肺炎病例累计确诊 17103 人. 韩国首例患者确诊于 1 月 20 日, 从 1 月下旬开始, 韩国政府将传染病预警级别上调为“警戒”. 从 2 月 4 日起, 韩国宣布入境限制, 即 14 天内有湖北旅居史者禁止入境. 截至 5 月 19 日, 韩国新冠肺炎病例累计确诊 11078 人. 北京市在疫情初始阶段控制严格, 但在 6 月份全国范围内几乎没有本地新增病例的情况出现第二轮暴发, 引起广泛关注. 截至 7 月 2 日, 北京市新冠肺炎病例累计确诊 923 人. 武汉市的首例新冠肺炎确诊患者于 2019 年 12 月 8 日出现症状并前往医院就诊. 在 2020 年 1 月, 疫情在武汉市迅速蔓延, 经我国政府采取多种紧急防控措施, 后期疫情的蔓延态势逐渐得到控制. 截至 5 月 13 日, 武汉市新冠肺炎病例累计确诊 50339 人. 在数据集中, 2 月 12 日的当日新增确诊病例突增 13436 例, 原因为当日临床诊断病例数被纳入确诊病例, 统计口径发生了改变, 使得确诊病例定义的范围扩大. 为消除此事件对疫情传播强度变化分析的影响, 本文以此时间点为界, 将疫情数据集分为 2019.12.8–2020.2.11 和 2020.2.13–2020.5.13 两个阶段分别进行分析. 从累计确诊病例数据集可知, 在第一阶段 (以下简称“武汉上”) 共有 19558 个确诊病例, 第二阶段 (以下简称“武汉下”) 共有 17718 个确诊病例. 本文将后期官方公布的核增或核减的病例数均纳入 2 月 12 日发生的改变中. 图 2 展示了各疫情数据集中的累计确诊病例总数随时间变化的趋势.

由于所收集的是每日新增确诊病例数, 而各个病例确诊的精确时间未知, 本文将当日每个新增确诊病例确诊时间均匀分布在当日 24 小时的时间区间内以模拟各个病例的确诊时间. 固定样条阶数为 $r = 4$ 阶, 内部结点为 $K_0 = 4$ 个. 表 2 展示了各疫情数据集的参数 μ 与 α 的估计值.

从表 2 可以看出反映由外来因素引发确诊病例的基准强度参数 μ 的估计, 日本较韩国更高. 这可能与早期日本发生钻石公主号事件有关, 而韩国对外来输入病例的控制较好. 在武汉市第一阶段初期, 确诊病例自发产生的强度较高, 在参数上体现为基准强度较大. 而在第二阶段, 政府已采取了严防输入病例与严格的进口冷链食品检疫等措

施，因此确诊病例的产生主要在于人与人之间的相互传染，所以基准强度非常小。从参数 α 的估计来看，相比于日本和武汉市，韩国和北京市的更久远的历史病例数均对当前病例的发生产生更强的促进作用。

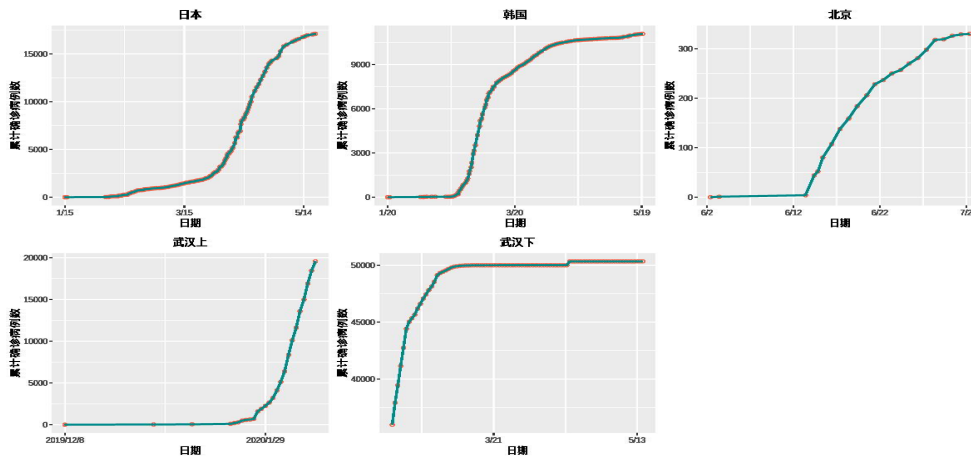


图 2 各疫情数据集的累计确诊病例总数随时间变化的曲线

表 2 各疫情数据集中 μ 与 α 的估计值

参数	日本	韩国	北京	武汉上	武汉下
μ	2.410	0.657	0.303	1.294	0.073
α	10.57	1.380	1.000	6.732	6.486

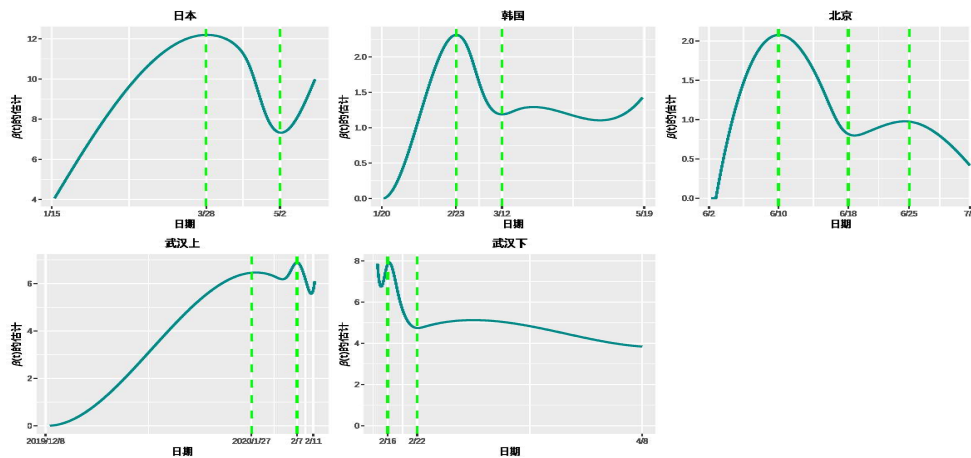


图 3 各疫情数据集的 $\beta(\cdot)$ 的函数曲线估计

各疫情数据集的 $\beta(\cdot)$ 的函数曲线估计如图 3 所示，可得到如下结论。在日本新冠肺炎疫情中，初期 2 月中旬发生钻石公主号事件，并在 2 月下旬呈现出疫情蔓延至医

务人员的态势,因此日本国内的新冠疫情可能会急速扩散.直到4月7日,首相安倍晋三发布紧急事态宣言,通过避免密闭空间,人员密集,密切接触等方式防止感染扩大,宣言实用时间为4月7日至5月6日(后延期一周).第一个拐点出现在3月28日,可能是由于政府发布的紧急事态宣言对疫情抑制起到积极作用.5月14日,日本政府会议决定从即日起在39个县接触紧急事态宣言,从图中可以看出,疫情传播强度出现反弹,说明政府的紧急事态宣言的发布和解除对疫情防控存在重要影响.

韩国疫情暴发在2月23日前后,韩国大邱市的一名61岁新天地教会女信徒成为超级传播者,导致大邱疫情暴发.仅在3天后,韩国就对大邱市进行最大限度的封锁.3月3日,韩国总统文在寅在首尔主持召开内阁会议,韩国政府进入24小时全面戒备状态.政策生效后,确诊病例人数保持陡降的态势.3月12日后 $\beta(\cdot)$ 的估计图像虽然没有呈现持续下降的趋势,但是保持在一个小范围内波动,说明韩国的疫情已得到有效控制.

6月北京市疫情暴发后,从6月13日起,新发地暂时关停市场,进行卫生整治和环境消杀.此外,北京市还采取了诸多措施,如扩大流行病学调查和检测的范围,对周边场所进行全面排查;暂停跨省旅游团,体育赛事,暂缓复课等.及时推出的防疫政策措施使得疫情很快得到控制,传播强度呈现逐渐降低的趋势,6月18日后每日新增的确诊病例数逐渐减少,在6月25日左右出现十分微弱的波动,之后疫情传播强度呈现下降的趋势,从6月28日起每日新增确诊病例仅为个位数,说明在非常短的时间内有效控制了北京市疫情的发展.

对于武汉市而言,从新冠肺炎传播强度变化的估计曲线可以看出,在第一阶段,拐点出现在2月7日左右,确诊人数在1月27日前不断攀升,之后出现小幅波动,直到2月7日达到极大值,在此之后出现下降态势.经过溯源,已确定第一位病人于2019年12月8日发病,于12月12日确诊.武汉市于2020年1月23日上午10点正式封城.1月24日,湖北省政府宣布启动重大突发公共卫生事件I级响应,各种聚集性活动一律取消.此后湖北省政府每日召开新冠肺炎疫情防控工作新闻发布会,公布最新的疫情防控政策与措施.正是这些科学合理的防控措施的及时出现,有效抑制了疫情的蔓延,使得确诊人数在1月27日后没有再持续上升的趋势.武汉大学医学部传染病教授,武汉大学中南医院感染科医生,中国艾滋病防治专家组成员桂希恩在1月26日接受媒体采访时表示预计正月十五前武汉市疫情可能出现拐点(https://www.guancha.cn/politics/2020_01_26_533212.shtml).2月7日是农历正月十四,与其预测相符.

而在第二阶段,第一个拐点出现在2月17日左右.2月11日,钟南山院士在接受新华社记者采访时表示目前疫情拐点还无法预测,但峰值应该在2月中下旬出现(http://www.xinhuanet.com/politics/2020-02/11/c_1125561318.htm).这一预测与本文所得的结果较为吻合,随着防疫政策的推进,疫情传播强度开始下降,在2月22日左右达到最小.然而随后,疫情传播强度又有小幅的上升.联系该时间点左右的事件可以得知,这一增长可能与监狱里疫情的蔓延有关.截止2月21日,湖北省监狱系统已发现271例新冠肺炎确诊病例,分别在两所监狱,湖北省武汉市女子监狱和沙洋汉津监狱,均为输入型病例.之后伴随着对监狱疫情控制的重视,推出了一系列的措施,使得疫情

传播强度在一个可控范围内波动, 并且波动幅度非常小. 直到 4 月 8 日, 武汉正式解封, 疫情传播强度一直呈现下降的趋势.

基于所得到的各参数的估计, 可以估计各疫情数据集的任何时间点的强度函数的大小. 图 4 揭示了各地疫情中变系数 Hawkes 过程的强度函数估计随新增病例数变化的规律. 虽然有部分日增陡升, 但对强度函数的影响有限, 说明这些病例一旦确诊即被控制住, 未能对其后确诊病例产生影响. 这对于武汉市第二阶段尤为明显, 说明了此时武汉市的防疫措施精准有效, 对遏制新冠肺炎的传播和最终战胜疫情起到重要的作用.

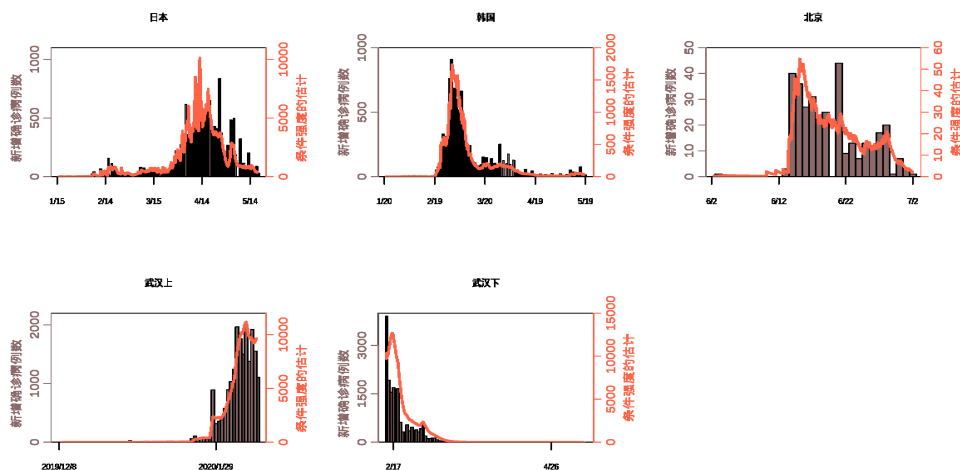


图 4 各地疫情中变系数 Hawkes 过程的强度函数随新增病例数变化的规律

5 小结

本文所提出的变系数 Hawkes 过程可以很好地揭示新型冠状病毒肺炎疫情的传播规律. 实际数据分析的结果表明政府推行了一系列强有力的防疫措施后, 疫情的传播能够得到有效控制. 此外, 在防疫过程中政府采取的措施旨在外防输入, 内防扩散, 这说明在模型 (2.2) 中, 与输入性病例有关的基准强度参数 μ , 也可以考虑为与时间相依的函数型参数 $\mu(t)$. 其次, 对于已有的单簇中下一个病例产生时间的分布 (亦即引发密度函数 $g(\cdot)$) 也可考虑除指数分布外的其它分布. 另外, 除了利用变系数的自激发性点过程考虑疫情传播强度在时间上的变化外, 还可以研究疫情传播在空间上的规律. 例如将不同地区的每日新增病例数据集视为动态的网络数据, 并运用互激发性点过程即多维 Hawkes 过程 [11,12] 研究不同地区之间的疫情互相影响和传播的规律 [13].

致谢 本文作者感谢两位审稿人仔细地评阅本文并提出宝贵的意见和建议.

参 考 文 献

- [1] Gao Y, Zhang Z, Yao W, Ying Q, Long C, Fu X. Forecasting the Cumulative Number of COVID-19 Deaths in China: a Boltzmann Function-Based Modeling Study. *Infection Control & Hospital Epidemiology*, 2020, 41: 841–843
- [2] Wu T J, Leung K, Leung M G. Nowcasting and Forecasting the Potential Domestic and International Spread of the 2019-nCoV Outbreak Originating in Wuhan, China: A Modelling Study. *Lancet*, 2020, 395: 689–697
- [3] Zhou F, Yu T, Du R, Fan G, et al. Clinical Course and Risk Factors for Mortality of Adult Inpatients with COVID-19 in Wuhan, China: A Retrospective Cohort Study. *Lancet*, 2020, 395: 1054–1062
- [4] Yu J, Song Y, Fei F, et al. Analysis of COVID-19 in Hubei Province Based on Machine Learning and Dynamics Model. *Journal of Biomedical Engineering Research*, 2020, 03: 294–299
- [5] Wang X, Yu Y, Hu Y, et al. COVID-19 Analysis and Forecast Based on Exponential Smoothing Model in Hubei Province. *Journal of Public Health and Preventive Medicine*, 2020: 1–4
- [6] Wang X, Tang S, Chen Y, et al. When will be the Resumption of Work in Wuhan and its Surrounding Areas During COVID-19 Epidemic? A data-driven Network Modeling Analysis. *Scientia Sinica Mathematica*, 2020, 50: 969–978
- [7] Hawkes A G. Spectra of Some Self-Exciting and Mutually Exciting Point Processes. *Biometrika*, 1971, 58: 83–90
- [8] Hawkes A G, Oakes D. A Cluster Process Representation of a Self-Exciting Process. *Journal of Applied Probability*, 1974, 11: 493–503
- [9] Kelly J D, Park J, Harrigan R J, Hoff N A, et al. Real-Time Predictions of the 2018–2019 Ebola Virus Disease Outbreak in the Democratic Republic of the Congo Using Hawkes Point Process Models. *Epidemics*, 2019, 28: 100354
- [10] Huang J, Wu Y. Change Point Detection of the Self-Exciting Point Process with Application to the Ebola Virus Disease. *Submitted*
- [11] Zhou K, Zha H, Song L. Learning Social Infectivity in Sparse Low-Rank Networks Using Multi-Dimensional Hawkes Processes. *AISTATS*, 2013, 31: 641–649
- [12] Gupta A, Farajtabar M, Dilkina B, and Zha H. Discrete Interventions in Hawkes Processes with Applications in Invasive Species Management. *In Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI 18)*, 2018, AAAI Press, 3385–3392
- [13] Fox E W, Short M B, Schoenberg F P, Coronges K D, Bertozzi A L. Modeling E-Mail Networks and Inferring Leadership Using Self-Exciting Point Processes. *Journal of the American Statistical Association*, 2016, 111: 564–584

Transmission Rate Analysis of COVID-19 Based on the Hawkes Process with Time-varying Coefficient

WU YUANSHAN[†]

(*School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan 430073, China*)

([†]*E-mail: wu@zuel.edu.cn*)

HUANG JIAYU

(*School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China*)

(*E-mail: jerry_huang@whu.edu.cn*)

YE ANJUN

(*School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan 430073, China*)

(*E-mail: bigyeanjun@163.com*)

Abstract Currently, the coronavirus disease (COVID-19) is becoming a pandemic around the world, threatening the human social and economic activities and life safety. It is scientifically interested to investigate the evolution process of COVID-19 with confirmed positive cases in specific region and time period. We propose a time-varying coefficient Hawkes process to study the trend of COVID-19 transmission over time and evaluate the effect of government intervention measures as well as some emergent events. Adopting the spline approximation technique, we develop a semiparametric estimating approach, whose finite-sample performance is assessed through simulation study. As an illustration, we apply the time-varying coefficient Hawkes process to examine the transmission dynamics of the pandemic in Japan, South Korea, Beijing and Wuhan, showing that the government intervention policy and emergent public health event can affect the evolution mechanism of COVID-19.

Key words point process; recurrent event; semiparametric method; spline functions; inflection point

MR(2000) Subject Classification 60G55; 60G35

Chinese Library Classification O211.65