

# 基于Markov链对上海市每日新增本土病例变化的预测

张 丁

上海工程技术大学管理学院, 上海

收稿日期: 2022年9月12日; 录用日期: 2022年10月2日; 发布日期: 2022年10月12日

---

## 摘 要

疫情防控是一项具有专业性、复杂性和全面性的系统性工作, 上海市每日新增本土病例变化对上海疫情防控有着十分重要的意义, 其单日新增变化趋势也是疫情防控的一项重要内容, 为后续疫情防控起到关键的引领作用。本文依据Markov链相关定义及定理, 引入Markov链以建立相应的模型, 对上海市每日新增本土病例变化状态进行了预测, 有利于动态改进疫情防控的流程, 提高管理效率。

## 关键词

Markov链, 单日新增, 转移矩阵, 预测

---

# The Prediction of the Daily Change of New Local Cases in Shanghai Based on Markov Chain

Ding Zhang

School of Management Studies, Shanghai University of Engineering Science, Shanghai

Received: Sep. 12<sup>th</sup>, 2022; accepted: Oct. 2<sup>nd</sup>, 2022; published: Oct. 12<sup>th</sup>, 2022

---

## Abstract

Epidemic prevention and control is a professional, complex and comprehensive systematic work, the daily changes in new local cases in Shanghai are of great significance to the prevention and control of the epidemic in Shanghai, and its single-day new change trend is also an important part of the epidemic prevention and control, playing a key leading role in the follow-up epidemic pre-

vention and control. Based on the relevant definitions and theorems of Markov chain, this paper introduces Markov chain to establish a corresponding model, and predicts the change status of new local cases in Shanghai every day, which is conducive to dynamically improving the process of epidemic prevention and control and improving management efficiency.

## Keywords

Markov Chain, Single-Day Additions, Transfer Matrix, Forecast

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

上海是全球最大的金融中心之一，有“东方巴黎”之称，常住人口高达 2400 多万，人口密度高度集中，一旦爆发大规模传染疾病，疫情防控难度较大，因此封控无疑是最有效的病毒切断措施之一，但上海作为中国的经济中心和消费中心，停摆将对中国经济造成严重影响，与此同时，在全球化时代，任何国家都不能独善其身，都将受到疫情的影响。从历史经验来看，人类周期性的受到传染病的挑战，其中大流行病是最严重的情况。“新型”冠状病毒是一种人类以前从未发现的新菌株，世界卫生组织于 2020 年 3 月 12 日宣布 COVID-19 (新型冠状病毒肺炎) 疫情形成全球大流行。2022 年已是新冠肺炎疫情在全球跌宕蔓延的第三个年头，新冠疫情对人类的生产生活产生重大影响。尽管我国本土 COVID-19 疫情已得到较好控制并实践获得丰富的经验，但由于病毒变异快，导致本次病毒的隐秘性极强，上海在 2022 年 3 月 1 日再次爆发疫情，使得我国未来疫情防控仍然面临较大潜在风险。因此，有必要对上海新冠肺炎新增本土病例变化展开深入研究，通过数据来分析未来疫情动向，一方面，能够研判我国各级政府面对突发性重大公共卫生事件的响应防控能力，另一方面，可以为我国未来疫情防控提供决策参考依据。

## 2. Markov 链概述

Markov 链是由俄罗斯科学家马尔可夫提出，用来描述一类重要的随机动力系统或过程的模型，在企业管管理、经济预测、水文气象、疾病预测等领域广泛应用。

孙桂娥利用 Markov 链预测了物流企业的应收账款风险，通过对应收账款回收周期来进行状态划分，建立 Markov 链模型，对坏账率和应收账款回收率进行了预测[1]；何冰洁建立回收处理的 Markov 链模型。对产品回收状态划分为 7 个状态，利用软件分析考察回收的效率、时间和成本，为管理者提供改善流程工艺提供了指引[2]；高发玲基于马尔科夫链对城乡人口流动状态进行了预测，为我国城镇化建设提供了重要的经验数据[3]。

此外，Markov 链预测有一个显著特点在于无后效性，无后效性指下时期状态只取决于本期状态和转移概率[4]，即要预测未来的状态，只需要知道当前的状态和概率就足够了，而不需要知道以往的情况，所以，通过 Markov 链预测可以利用系统现状及其发展动向去预测该系统的未来状况。基于 Markov 链模型的诸多优良特性，本文也将继续延续此话题，建立上海市每日新增本土病例变化的数学模型，进行定量分析，希望能为上海疫情防控提供数据参考。

## 3. 马尔科夫链模型的建立

构建马尔科夫链模型的目标是为了体现上海市每日新增本土病例变化的特征，即在不同的时间序列

阶段中会呈现出不同的新增数量的病例变化,利用 Markov 链模型就是为了预测不同的时间序列阶段新增数量的病例变化。首先,需要厘清上海市每日新增本土病例变化主要取决于何种因素以及在构建模型时上海市每日新增本土病例变化所处的状态,也即初始状态。把上海市每日新增本土病例变化的时间序列看作马尔科夫链,就可以根据当前的状态对未来的状态进行预测,通过数据分析之后,进而可以采取相应的策略,这就是利用马尔科夫链进行上海市每日新增本土病例分析的基本思想[5]。通过获取上海单日新增本土病例人数的历史资料,得到连续两天内前天处在状态  $i$  区,第二天处在  $j$  区的比率  $P_{i,j}$ ,构造出一步转移矩阵  $P_1 = P_{i,j}$ ,  $k$  步转移概率矩阵  $P_k$  为:

$$P_k = P_1^k$$

记向量  $P(t) = \{P_1(t), P_2(t), \dots, P_n(t)\}$ , 其中  $P_i(t)$  表示在第  $t$  天上海单日新增本土病例处于第  $(i \in E)$  的绝对概率。 $E$  即: 如果预测对象有  $E_i$ , 有  $(1, 2, \dots, n)$  个状态, 某一状态出现的次数为  $M_i$  则  $E_i$  出现的频率  $F_i = M_i/N$ 。  $F_i = M_i/N$  上海单日新增本土病例第  $t+k$  个时间段  $(k \in T)$  的绝对概率向量:

$$P(t+k) = P(t)P(k) = P(t)P_1^k$$

因此, 当我们知道当前的初始概率向量所处的状态后, 就可以预测未来任意一个时间的状态此外, 容易看出上海单日新增本土病例人数的马氏链具有遍历性, 也就是说, 在若干期后, 上海单日新增本土病例人数最终在各个期间的概率分布都是一个平稳值。

#### 4. 应用实例分析

现以上海市每日新增本土病例时间序列为例(见表 1), 应用马尔科夫链对每日新增本土病例分别进行中短期和长期预测分析, 从而确定策略, 将时间序列的单位以日记。

**Table 1.** Shanghai has daily new local cases from 2022.5.5 to 2022.5.25

**表 1.** 上海市每日新增本土病例 2022.5.5~2022.5.25 资料

日期	单日新增本土病例	日期	单日新增本土病例
2022.5.5	245	2022.5.16	77
2022.5.6	253	2022.5.17	96
2022.5.7	215	2022.5.18	82
2022.5.8	322	2022.5.19	88
2022.5.9	234	2022.5.20	84
2022.5.10	228	2022.5.21	52
2022.5.11	144	2022.5.22	55
2022.5.12	227	2022.5.23	58
2022.5.13	194	2022.5.24	44
2022.5.14	166	2022.5.25	48
2022.5.15	69		

将这 21 日上海市每日新增本土病例划分为 4 个区间, 得到区间状态为: 1 (每日新增本土病例 113 以下), 2 (每日新增本土病例 113~182), 3 (每日新增本土病例 182~252), 4 (每日新增本土病例 252 及以上)。综合这些资料得到这 21 天的海市每日新增本土病例状态转移情况如表 2 [6]。

**Table 2.** Shanghai 2022.5.5~2022.5.25 daily new local case status transfer in 21 days  
**表 2.** 上海市 2022.5.5~2022.5.25 日 21 天内每日新增本土病例状态转移情况

状态	次日状态 1	次日状态 2	次日状态 3	次日状态 4
当日状态 1	10	0	0	0
当日状态 2	1	0	1	0
当日状态 3	0	2	2	2
当日状态 4	0	0	2	0

由此计算出各状态之间的转移概率和转移概率矩阵:

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

根据表 1 显示, 2022.5.25 上海市每日新增本土病例人数的状态为 1, 可以认为初始状态向量  $S^{(0)} = (1, 0, 0, 0)$ , 可根据一步转移矩阵预测 2022.5.26 年上海市每日新增本土病例各个状态的概率:

$$S^{(1)} = S^{(0)}P = (1, 0, 0, 0) \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix} = (1, 0, 0, 0)$$

得到 2022.5.26 上海市每日新增本土病例人数位于 1 状态的概率为 1, 位于 2 状态的概率为 0, 位于 3 状态的概率为 0, 位于 4 状态的概率为 0。2022.5.26 上海新增本土病例人数为 45 人, 位于 [0, 113), 属于状态 1, 与实际情况相符。

$$S^{(2)} = S^{(0)}P^2 = S^{(1)}P = (1, 0, 0, 0) \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix} = (1, 0, 0, 0)$$

得到 2022.5.27 海市每日新增本土病例位于 1 状态的概率为 1, 位于 2 状态的概率为 0, 位于 3 状态的概率为 0, 位于 4 状态的概率为 0。2022.5.27 上海新增本土病例人数为 39 人, 位于 [0, 113), 属于状态 1, 与实际情况相符。

利用 matlab 进行计算, 得出:

$$S^{(90)} = S^{(n)} = S^{(0)}P^n = (1, 0, 0, 0) \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix} = (1, 0, 0, 0)$$

即 2022.5.25 之后上海市每日新增本土病例人数位于 1 状态的概率为 1, 但是若要在新增人数上达到平稳, 要在 90 日后才能实现新增人数上的平稳, 此时, 每天新增人数为 0。

上海市每日新增本土病例人数为齐次马尔科夫链, 则  $P_n = P_1 P_1^{n-1} = P_1^n (n \geq 1)$ , 且若它的状态空间  $E$  是有限的, 使得  $\lim_{n \rightarrow \infty} P_{i,j}(n) \pi(j)$ , 马尔科夫链具有遍历性, 且  $\pi(j) > 0$ ,  $\sum_j \pi(j) = 1$  的唯一解, 即经

过一段时间之后，马尔科夫链会达到平稳状态。又由上述分析可知：

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

基于此，联立方程组：

$$\begin{cases} \pi(1) = \pi(1) \\ 1/2\pi(1) + 1/2\pi(3) = \pi(2) \\ 1/3\pi(2) + 1/3\pi(3) + 1/3\pi(4) = \pi(3) \\ \pi(3) = \pi(4) \\ \pi(1) + \pi(2) + \pi(3) + \pi(4) = 1 \end{cases}$$

解得：

$$\begin{cases} \pi(1) = 1/4 \\ \pi(2) = 1/4 \\ \pi(3) = 1/4 \\ \pi(4) = 1/4 \end{cases}$$

即为较长时间后上海市每日新增本土病例人数处于各区间的平稳分布，且又因为  $\pi(1) = \pi(2) = \pi(3) = \pi(4)$ ，上海市每日新增本土病例人数又为均匀平稳分布。

## 5. 结论与启示

新冠肺炎作为全球性流行传染病给世界造成重大影响，严重危害人类生命健康安全，本文依据 Markov 链相关定义及定理，引入 Markov 链以建立相应的模型，对上海市每日新增本土病例变化状态进行了预测，得到如下结论：1) 不管 2022 年的上海市每日新增本土病例怎样变化，经过数月后都会趋于平稳；2) 2022.5.25 之后 90 日才能实现新增人数上的平稳。与实际值对比发现，Markov 链预测上海市每日新增本土病例的结果是比较满意的。值得注意的是，马尔科夫链的结果并不是为我们提供一个准确又具体的数值，而是基于其无后效性，根据最初的状态划分，对未来可能的状态趋势进行预测，其结果代表的只是一种状态趋势。在实际过程中，诸如新冠肺炎等大型传染病的传播往往受到许多因素的影响，因而对状态趋势的预测也并非一成不变，所谓的平稳也是动态平稳，未来仍可能受到其他因素的干扰而发生改变，比如说小区的执行力，居民的自觉性等[7]。结合我国的出行习惯，虽然马尔科夫链的结果 2022.5.25 之后 90 日才实现新增人数上的平稳，但在每年 9 月份，全国的大、中、小学将面临开学，人口流动性将会大大提高，届时马尔科夫链的状态可能仍会发生转移。此外，上海每日新增本土病例并无定律，没有什么方法可以对每日新增本土病例进行准确预测。

不过，即便如此，Markov 链预测疫情的传播仍然具有一定的借鉴作用，只要根据各地区的不同特点，进行具体分析，充分利用大数据进行检测，仍然可以得到与真实状态接近的结果。本文根据马尔科夫链的预测结果对上海市疫情防控提出以下建议：1) 在疫情爆发的高峰阶段，首先要摸清疫情源头，从根本上阻断疫情的传播，采取强有力的隔离措施，避免人群积聚，对密接人群要实行严格的隔离观察，防止二次传播；2) 及早做好疫情防控的风险评估工作，做好防控预案，加强群众心理建设，避免引起群众恐慌；3) 保证数据的真实性、完整性和准确性是做好疫情管控的重中之重，还要保证平台数据的同步性、及时性和透明性，避免误导民众，有助于及时调控资源和医疗基金。

## 参考文献

- [1] 孙桂娥. 基于 Markov 链的物流运输企业应收账款风险分析[J]. 重庆交通大学学报(社会科学版), 2010, 10(4): 18-20.
- [2] 何冰洁, 岳玉静. 基于 Markov 链的产品回收处理预测[J]. 上海工程技术大学学报, 2010, 24(2): 105-109.
- [3] 高发玲. 基于 Markov 链预测研究城乡人口流动状态[J]. 辽宁大学学报(自然科学版), 2017, 44(2): 118-120.
- [4] 李美桂, 李壮壮. 基于 Markov 链在 GDP 增长率中的预测[J]. 产业与科技论坛, 2010, 9(4): 109-110.
- [5] 叶焯, 劳飞翔, 韦斯亮, 叶力, 周红霞. 1950-2014 年广西横县疟疾疫情流行病学分析及 Markov 模型预测[J]. 微生物学免疫学进展, 2019, 47(1): 54-59.
- [6] 刘次华. 随机过程[M]. 第 2 版. 武汉: 华中科技大学出版社, 2001.
- [7] 万光辉, 石裕祥. Markov Chain 在流脑疫情预测中的应用[J]. 数理医药学杂志, 1993, 6(2): 54-55.