

马尔科夫链和流行病学：流行病在人群中的传播

3I005

2022年11月21日

这个项目的目的是操纵马尔科夫链来研究流行病在人群中的传播。你的输出将是一个笔记本，希望对代码进行**评论**并对**结果进行解释**。推荐使用numpy、random和matplotlib软件包。

我们将研究由3种类型的个体组成的人群。每个人都处于三种**状态**之一：健康的**S**，感染的**I**或治愈的**R**。

1 从数据中学习模型参数

在我们的模型中，我们将考虑在每个时间：

- 每个健康人都可以保持健康或被感染
- 每个受感染的人都可以保持感染或变得痊愈
- 每个被治愈的人都能保持痊愈

而从一个状态到另一个状态的概率**只取决于前一个状态**。

我们有一个观察序列，我们想学习马尔科夫链的参数来模拟产生该序列的基本过程。我们对一个人进行了**10天**的跟踪，以确定在每个时间段，这个人处于哪种状态。我们得到了以下观察序列。**s, s, s, i, i, i, i, i, i, r**。

1. 根据这一连串的观察，估计各状态之间的过渡概率，并绘制出过渡概率矩阵。

然后我们对**5000**个个体的群体进行了**200**天的监测。要读取数据，请使用
`np.loadtxt('data_ex01.txt')`。健康的人得**0**分，感染的人得**1**分，治愈的人得**2**分。

1. 读取数据
2. 估计各状态之间的过渡概率并构建过渡概率矩阵（提示检查你的结果：矩阵的第一行是[0.9308, 0.0691, 0.]）。

2 第一个模型的描述

一个人在时间**t**处于这三种状态之一的概率，只取决于他在时间**t-1**所处的状态。

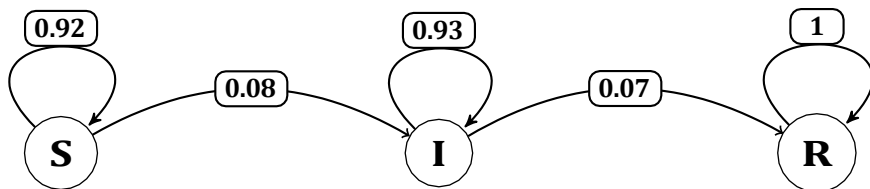


图1 - 显示3个模型状态和过渡概率的图表 1

一个处于健康状态的个体有0.92的概率保持健康，有0.08的概率成为感染者。如果个人被感染了，他可以以0.93的概率保持感染，以0.07的概率被治愈。

1. 从过渡图中，创建过渡矩阵 A ，该矩阵包含不同状态之间的过渡概率。创建一个函数来检查一个矩阵是否是随机的。

在时间 $t=0$ 时，一个人有0.9的概率是健康的，0.1的概率是被感染的。

2. 创建 π_0 ，即初始概率分布。

我们将研究健康、感染和治愈的个体数量随时间的演变，首先研究理论分布，然后在模拟上观察分布。

分布 π_t

让 π_t 是时间 t 的概率向量。

1. 使用 π_0 和 A ，给出一个人在时间上健康、被感染或治愈的概率 $t=1$ （先用手计算一下）。
2. 给出一个人在时间 $t=2$ 时健康、被感染或治愈的概率（先用手计算）。
3. 同样地，对于1到200之间的每个时间 t ，计算出数字的理论分布，在每个州（回顾 $\pi_{t+1} = \pi_t A$ ）
4. 绘制处于每种状态的概率与时间的关系图（+描述一下你观察到的情况）。

随机抽出的状态

你将使用这个马尔可夫链生成一个大小为 T 的序列。为了产生一个随机序列，随机选择一个初始状态（使用 π_0 ）；然后按照过渡概率（=过渡矩阵 A ）来选择后续状态。

你可以取 $T=150$ 。

建立一个人口模型

你已经为一个人生成了一个状态序列。现在你将为200个个体的群体生成一组序列。

1. 在每个时间 t ，计算人口中健康、感染和治愈的个体数量，并显示三种状态下的个体数量随时间的变化。
2. 显示健康感染者和治愈者的百分比是时间的函数。
3. 当 t 很大时，健康人、感染者和治愈者的比例是多少？
4. 用不同规模的种群重复前面的问题，如5个个体和5000个个体。

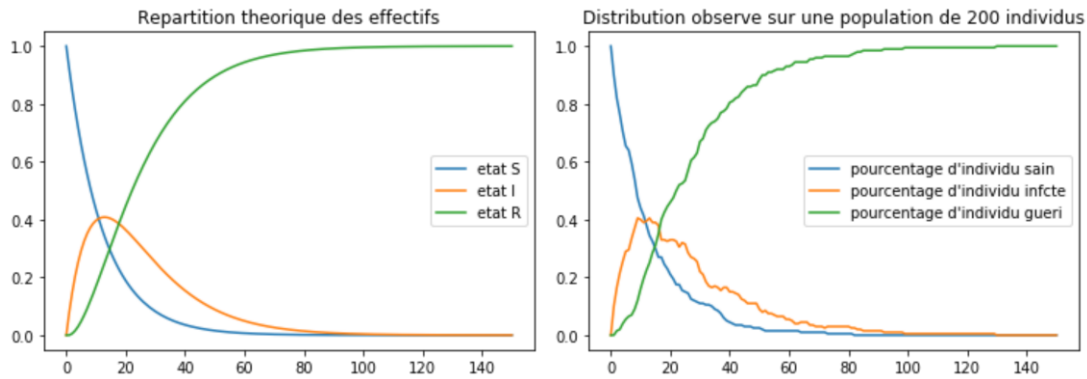


图2--理论上和观察到的每个状态下的个体百分比是时间的函数

感染的时间长度

1. 根据模拟结果，估计一个 I 序列的平均长度。
2. 显示观察到的感染长度的分布（也许为了表示分布，做一个直方图是个好主意）。
3. 计算一序列 I 的理论长度（你可以使用几何定律的期望值并说明理由）。
4. 将估计的长度与理论的长度进行比较。
5. 显示感染长度的理论分布。

围绕这第一个模型进行的小修改

你现在可以修改模型来研究不同的情况，例如通过改变

1. 人口规模
2. 的初始概率分布。
3. 过渡概率。

3 第二个模型的描述

现在我们将考虑第二个模型，即康复的个体可以以0.02的概率重新变得健康，他们可以失去对疾病的免疫力。

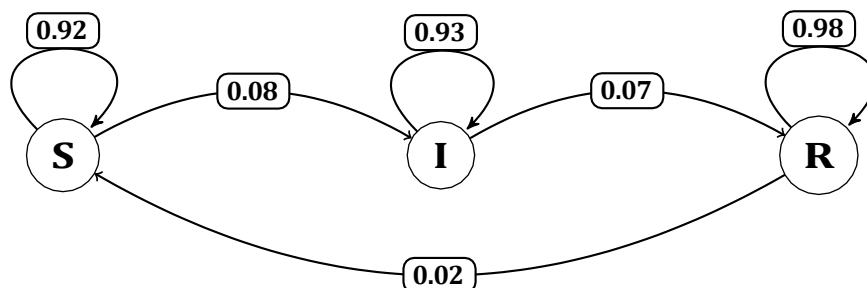


图3 - 显示3个模型状态和过渡概率的图表

模型的分析

1. 这个过程可以用马尔科夫链来模拟吗？
2. 给出新的过渡矩阵
3. 这个马尔可夫链中的状态的性质是什么？它是定期的吗？它是不可还原的吗？
4. 计算矩阵 $A \times A$ 。它对应的是什么？它是随机的吗？ A 的问题相同³和 A^4 。
5. 进行新的模拟。人口是如何变化的？
6. 用不同的初始概率分布重复模拟（例如，如果在时间 $t=0$ 时，我们有90%的感染者和10%的健康者）。探索其他的初始化，至于每个问题，请评论你的观察。
7. 计算静止的概率分布，并将这一结果与模拟结果进行比较，以获得规模相当大。

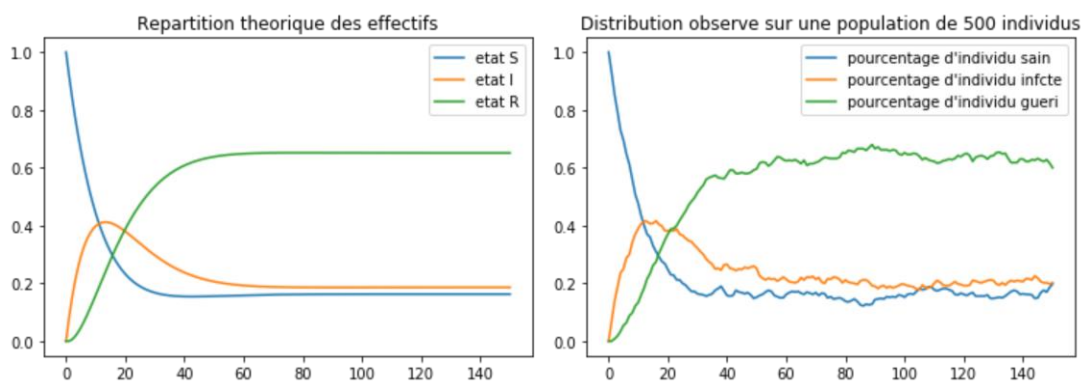


图4--理论上和观察到的每个状态下的个体百分比是时间的函数

豁免权的长度

一个患病的人在多长时间内保持对疾病的保护是值得怀疑的。

1. 根据模拟结果，估计 R 的序列的平均长度
2. 计算一个 R 的序列的理论长度
3. 显示豁免权长度的理论和观察分布。

修改模型

1. 如果你改变了一个健康个体被感染的概率，那么流行病是如何演变的？什么是新的均衡分布？
2. 同样的问题，如果你改变一个被治愈的个体重新变得健康的概率。

4 遏制

可以想象，如果社会疏远措施到位，成为感染者的概率就会变成零。

1. 我们将交替进行不疏导和疏导期。
 1. 在非遏制期，我们使用之前练习的过渡矩阵。
 2. 在遏制期，健康个体被感染的概率变为零
- 用前面练习中的过渡矩阵开始模拟。可以假设，在最初的时候，所有的人都是健康的。
- 当人口中有25%的感染者时，我们进入了一个疏远期
- 感染者的数量将减少。当感染者少于10%时，遏制措施就会解除。
2. 对一个足够大的种群进行模拟，在每个时间 t 绘制个体数量的演变（你应该看到“波浪”），并注意遏制和解禁的时间。
3. 需要多少个封闭/解封装置？

5 讨论

1. 你能对所使用的模型提出什么批评意见？
2. 可选：提出改进建议，并在可能的情况下实施。