

法国冠状病毒流行开始的数学模型

Nicolas Bacaër

Institut de Recherche pour le Développement
复杂系统的数学和计算机建模单位

, 法国

巴黎莱斯科德利耶尔 nicolas.bacaer@ird.fr

<https://hal.archives-ouvertes.fr/hal-02509142>

总结

我们将2020年法国新型冠状病毒的流行数据拟合为SEIR类型的流行模型。如果从特定日期起将接触减少为零 T 接近流行病的开始, 流行病的最终规模接近于病例总数的乘积 $R(T)$ 在此日期的可复制性 \mathcal{R}_0 流行病。更一般而言, 如果联系是随时间推移而划分的 T 由 $q > 1$ 这样 $\mathcal{R}_0/q < 1$, 那么该流行病的最终规模接近 $R(T) \mathcal{R}_0 (1 - 1/q)/(1 - \mathcal{R}_0/q)$ 。根据我们的估计, $\mathcal{R}_0 \simeq 2,3$ 在法国。

1.模型

图1^a显示了2020年2月25日至3月29日在法国检测到的冠状病毒病例的累计数量。我们必须区分3月15日是突然采取严厉措施制止该流行病的日期(学校停课, 餐厅等)。在这三个日期中, 累计案件数从13起增加到5 427 然后 40 174。图1^b显示的数据相同, 纵坐标为对数坐标, 以及线性回归线。分为三个阶段: 第一阶段一直持续到3月6日, 增长迅速但相当不规律。在第二年(截止到3月15日), 增长略慢但稳定。第三, 从3月16日开始, 增长缓慢而稳定。如果我们在从2月25日到3月15日的前两个时间段调整一条线, 则会发现案件数量随着 $e^{\lambda t}$ 增长速度 $\lambda \simeq 0,31$ 每天(右红色)。倍增时间是 $\log(2)/\lambda \simeq 2,2$ 天。另一方面, 如果我们将自己限制在第二个时期, 并且数据在对数尺度上特别好地对齐, 则可以得出 $\lambda \simeq 0,225$ 每天加倍的时间 3,1 天(蓝色右侧)。由于流行开始时的数据受到大量新输入病例的干扰, 因此第二个估计可能更可靠; 这是我们将在下面的干预前阶段使用的那个。至于第三期, 我们发现增长率为0,141 每天加倍的时间 4,9 天。

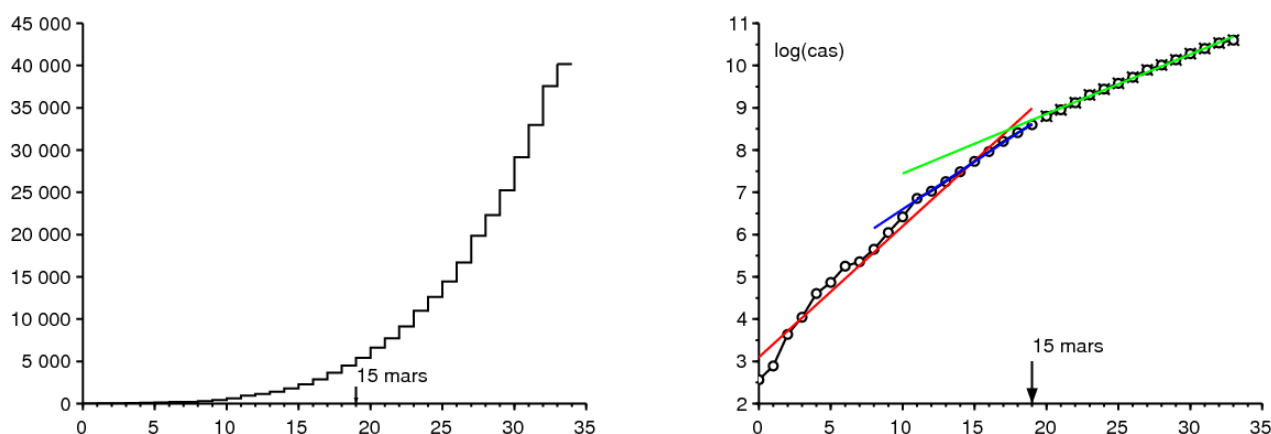


图1. a) 根据法国公共公报, 法国在2020年2月25日至3月29日期间累计发现病例数。b) 病例数和线性回归线的自然对数。

让我们为这种流行病提出一个数学模型。让我们根据经典的SEIR模型将法国人口分为四个部分(例如, 参见[3, p. 61]): 可能被感染了 (S), 已感染但未传染 (E), 没有保护的传染性 (I) 并从传动链中移除 (R)。最后一个隔间将传染性但受限, 治愈和死者的人群聚集在一起。显然, 我们可以无限地优化此模型以使其更逼真, 但是我们尝试了尽可能地限制未知参数的数量。我们的目标还在于获得比实际更具理论性的结果。

注意事项 N 总人口 a 有效接触率 b 感染者的传染率和 c 传染病隔离率:

$$\frac{dS}{dt} = -aS \frac{I}{N} \quad (1)$$

$$\frac{dE}{dt} = aS \frac{I}{N} - bE \quad (2)$$

$$\frac{dI}{dt} = bE - cI \quad (3)$$

$$\frac{dR}{dt} = cI, \quad (4)$$

与 $N = S(t) + E(t) + I(t) + R(t)$ 。为了与数据建立链接，我们可以认为 $R(t)$ 对应于目前已知的累计病例数 t 。

流行开始时，病例数与总人数相比仍然很小，因此 $S(t) \simeq N$ ，导致线性化

$$\frac{dE}{dt} \simeq aI - bE, \quad \frac{dI}{dt} \simeq bE - cI.$$

因此，这种流行趋势随着 $e^{\lambda t}$ ，在哪里 λ 是矩阵的最大特征值

$$\begin{pmatrix} -b & a \\ b & -c \end{pmatrix}. \quad (5)$$

特征多项式为

$$\lambda^2 + (b+c)\lambda + b(c-a) = 0. \quad (6)$$

所以

$$\lambda = \frac{-(b+c) + \sqrt{(b+c)^2 - 4b(c-a)}}{2} = \frac{-(b+c) + \sqrt{(b-c)^2 + 4ab}}{2}. \quad (7)$$

车厢中的平均时长 E ，这值得 $1/b$ 我们将其吸收到潜伏期的时间约为4天[4]。因此拿 $b = 0,25$ 每天。车厢中的平均时长 I 隔离之前，这值得 $1/c$ 这种病毒很难估计，因为它不仅取决于病毒的生物学特征，还取决于分离病例的及时性，因国家而异。假设距法国流行病大约1天，当时居民已经充分意识到该流行病的存在；患者很快被隔离。所以 $c = 1$ 每天。我们从公式 (7) 推导出

$$a = \frac{(2\lambda + b + c)^2 - (b - c)^2}{4b} = (\lambda + c) \left(1 + \frac{\lambda}{b}\right). \quad (8)$$

这将允许通过数字计算有效接触率 a 从观察到的增长率 λ 。

想象一下，公共卫生措施可以将有效接触率除以数字 k 大于1。至少必须值多少钱 k 制止流行病？这个值 k ，传统上 \mathcal{R}_0 跟随洛特卡并被她称为“可复制性”[6, p. 1. 102]，只需注意 a 被替换为 $a' = a/\mathcal{R}_0$ ，该流行病的新增长率 λ' 必须为零，根据等式 (6) 得出 $c - a/\mathcal{R}_0 = 0$ 和

$$\mathcal{R}_0 = \frac{a}{c} = \left(1 + \frac{\lambda}{b}\right) \left(1 + \frac{\lambda}{c}\right) \simeq 2,33$$

如果我们使用数值 ($\lambda \simeq 0,225$ 每天) 由图1的流行曲线建议。查看参数的不确定性 b 和 c ，这只能是一个近似值。

以一种更技术性的方式 (例如参见[7])，人们可能已经注意到 \mathcal{R}_0 也是矩阵的光谱半径

$$\begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix} \begin{pmatrix} b & 0 \\ -b & c \end{pmatrix}^{-1}.$$

让我们回到SEIR模型 (1) - (4)。让我们记住在完全没有干预的情况下如何确定流行病的最终规模；这是对SIR模型所用方法的简单改编 (例如，参见[5, 第76页])。式 (1) 表明

$$\frac{d}{dt} \log S = -\frac{a}{N} I(t).$$

因此，通过整合 $t = 0$ 在 $t = +\infty$,

$$\log S(\infty) - \log S(0) = -\frac{a}{N} \int_0^{\infty} I(t) dt. \quad (9)$$

流行开始时，没人在隔间里 R ，因此 $R(0) = 0$ 。等式 (4) 表明

$$R(\infty) = c \int_0^{\infty} I(t) dt, \quad (10)$$

在哪里 $R(\infty)$ 指定极限时 $t \rightarrow +\infty$ 功能的 $R(t)$ ，并以 N 。此外，我们一直都有 $S(t) + E(t) + I(t) + R(t) = N$ 。什么时候 $t \rightarrow +\infty$ ，这种流行病最终停止了， $E(t)$ 和 $I(t)$ 趋向于0。因此，在极限处，只有那些逃脱了该流行病的人和那些被感染但进入了该隔离区的人 R ：

$$S(\infty) + R(\infty) = N. \quad (11)$$

通过组合 (9)，(10) 和 (11)，我们看到

$$N - R(\infty) = S(0) \exp\left(-\frac{a}{c} \frac{R(\infty)}{N}\right).$$

流行开始时，人口中只有少数感染者，因此 $S(0) \simeq N$ 。流行的最终规模方程可以写成

$$1 - \frac{R(\infty)}{N} \simeq \exp\left(-\mathcal{R}_0 \frac{R(\infty)}{N}\right), \quad (12)$$

碰巧具有与SIR模型相同的形状[5]。用 $\mathcal{R}_0 \simeq 2,33$ ，我们找到了数字 $R(\infty)/N \simeq 87\%$ 。

2.第二阶段的大力干预

想象一下在某个日期 T ，采取了严厉的措施，以使新的有效接触率在有 $R(T)$ 累积案例。例如，截至3月15日，与学校和公共场所有关的措施生效之日，累计发生了5427起案件。然后，我们可以预测在这些假设下新的最终流行病的规模是多少？

在指数阶段结束时 $t \leq T$ 而案件总数仍只占总人口的一小部分，

$$E(t) \simeq u e^{\lambda t}, \quad I(t) \simeq v e^{\lambda t}, \quad R(t) \simeq w e^{\lambda t},$$

在哪里 (u, v) 是与最大特征值相关的特征向量 λ (5) 的矩阵。所以 $-bu + av = \lambda u$ 。通过公式 (8)，我们发现

$$u = \frac{av}{\lambda + b} = \frac{\lambda + c}{b} v.$$

作为 $dR/dt \simeq \lambda R$ 为 $t < T$ 什么时候 t 不太接近0，我们有

$$I(T) = \frac{1}{c} \frac{dR}{dt}(T) \simeq \frac{\lambda}{c} R(T).$$

但是 $I(T) \simeq v e^{\lambda T}$ ，因此

$$E(T) \simeq u e^{\lambda T} = \frac{\lambda + c}{b} v e^{\lambda T} \simeq \frac{\lambda + c}{b} I(T) \simeq \left(\frac{\lambda^2}{bc} + \frac{\lambda}{b}\right) R(T).$$

联系人应该减少到零，我们有 $t > T$

$$\frac{dS}{dt} = 0, \quad \frac{dE}{dt} = -bE, \quad \frac{dI}{dt} = bE - cI, \quad \frac{dR}{dt} = cI. \quad (13)$$

无需解决此系统，很明显，该流行病的最终规模将是 $R(\infty) = R(T) + E(T) + I(T)$ ，因为有 $E(T) + I(T)$ 尚未在车厢中的感染者 R 当时 T 。所以

$$R(\infty) \simeq R(T) \left(1 + \frac{\lambda^2}{bc} + \frac{\lambda}{b} + \frac{\lambda}{c}\right) = R(T) \left(1 + \frac{\lambda}{b}\right) \left(1 + \frac{\lambda}{c}\right) = \mathcal{R}_0 R(T).$$

因此，如果从接近流行病开始的某个日期开始将接触减少为零（足够接近以使线性逼近仍然有效，但对于接近线性化系统并没有足够时间有时间收敛）朝向与第一个特征值相关的特征向量），那么该流行病的最终规模就接近于该日的病例

累计数量乘以可再现性所获得的最终规模 \mathcal{R}_0 流行病。对于SIR模型，以相同的方式获得了相似的结果。在附录中，我们注意到，这不再是 \mathcal{R}_0 谁决定比例 $R(\infty)/R(T)$ 在传染期不是按指数分布，而是更复杂的表达式的模型中。

让我们通过人口统计学中人口“增长潜力”的概念进行类比[8, p. 1]。176]。如果生育率在该瞬间突然除以再现性，则它是最终固定人口与某一瞬间人口之间的比率 \mathcal{R}_0 ，因此人口的零增长率渐近线。在我们的计算中，通过假设初始种群在Lotka的意义上是“稳定的”（也就是说，由第一个特征向量给出的），Keyfitz获得了一个相对简单的公式来计算增加，这个公式也涉及 \mathcal{R}_0 尽管以比我们的SEIR模型更复杂的方式进行[8, p. 1。179]。

另请注意， $R(T) + E(T) + I(T)$ 从数据 $R(T)$ 只有一个类似于在艾滋病流行初期就出现的问题，即从报告的艾滋病病例数中估算艾滋病病毒阳性人数。

让我们回到我们的模型。用 $R(T) = 5427$ 和 $\mathcal{R}_0 \simeq 2,33$ ，它给 $R(\infty) \simeq 12600$ 。再次强调参数周围的不确定性 b 和 c ，这体现在 $R(\infty)$ 。

图2^a说明了此两阶段模型。我们拿了 $N = 65 \times 10^6$ （法国的人口）， $S(0) = N - 1$ ， $E(0) = 1$ ， $I(0) = 0$ 和 $R(0) = 0$ 。参数 a 由公式 (8) 给出 $\lambda = 0,225$ 每天。我们拿了 $T = 40,1$ 天 $R(T) \simeq 5415$ 与数据非常接近 54273月15日。通过继续进行比图中更长的模拟，我们在数值上发现 $R(\infty)/R(T) \simeq 2,33 \simeq \mathcal{R}_0$ 。

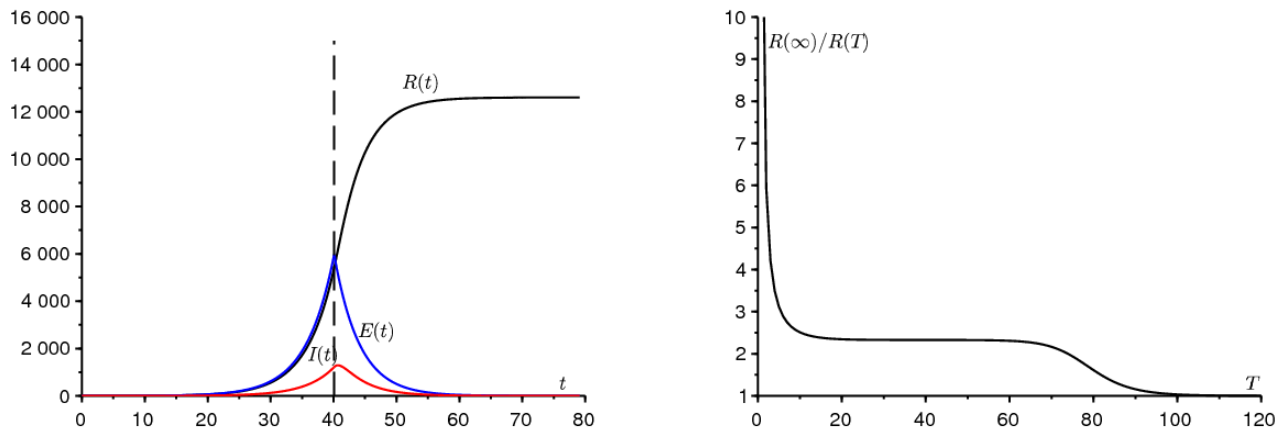


图2. a) 两相模型的仿真示例。b) 报告 $R(\infty)/R(T)$ 取决于 T 。

图2^b 显示报告如何 $R(\infty)/R(T)$ 因时间而异 T 接触率降低到零。我们可以清楚地观察到这个比率接近 \mathcal{R}_0 。什么时候 $T \rightarrow 0$ ，我们有 $R(T) \rightarrow 0$ 和 $R(\infty) \rightarrow E(0) + I(0) = 1$ ，因此报告 $R(\infty)/R(T)$ 趋于无穷大。该报告越来越接近 \mathcal{R}_0 什么时候 T 是矩阵 (5) 的两个特征值之差的倒数。相反 $T \rightarrow \infty$ ，那么干预来不及了；流行病已经过去， $R(\infty)/R(T) \rightarrow 1$ 。预计架子的宽度 $R(\infty)/R(T)$ 接近 \mathcal{R}_0 增长为常数乘以 $\log(N)$ 什么时候 $N \rightarrow +\infty$ ，因为这是具有常数系数的SIR模型中直到流行病高峰的时间行为[2]。

如果我们想知道何时会出现新病例高峰，这对于医院的交通拥挤问题可能很重要，则我们必须解决该系统 (13)。我们发现 $E(t) = E(T) e^{-b(t-T)}$ 和

$$I(t) = I(T) e^{-c(t-T)} + bE(T) \frac{e^{-b(t-T)} - e^{-c(t-T)}}{c - b}.$$

用图2的数值 a ，我们有 $E(T) \simeq 5970$ ， $I(T) \simeq 1218$ 和 $I(t)$ 在联系率降低到0后0.6天取消。对于非零联系率，延迟显然会更长。

3.概括

实际上，有效接触率肯定不完全为零 $t > T$ 。获得的价值 $R(\infty)$ 但是可以将其视为实际值的下限，因为可以肯定的是，对于非零接触，流行的最终规模将大于零接触， $t > T$ 。在这个问题上，应该记住，接触率可变的SIR或SEIR类型的流行病模型不是“单调的”，从某种意义上讲，接触率的降低有时会导致更大的最终流行病规模。 [1]。

现在考虑将接触率不降低到0而是简单地除以数字的情况 $q > 1$ 这样 $\mathcal{R}_0/q < 1$, 而 $\mathcal{R}_0 > 1$ 。减少到0对应于临界情况, 其中 q 趋于无穷大。在这种情况下, 我们有 $t > T$,

$$\frac{dS}{dt} = -\frac{a}{q} S \frac{I}{N}, \quad \frac{dE}{dt} = \frac{a}{q} S \frac{I}{N} - bE, \quad \frac{dI}{dt} = bE - cI, \quad \frac{dR}{dt} = cI.$$

根据与第1节相同的理由, 我们有 $t > T$

$$\frac{1}{S} \frac{dS}{dt} = -\frac{a}{q c N} \frac{dR}{dt}.$$

通过整合 $t = T$ 和 $t = +\infty$, 我们推断

$$\log \frac{S(\infty)}{S(T)} = -\frac{\mathcal{R}_0}{q N} (R(\infty) - R(T)).$$

假设当时在第2节中 T 既不能太小也不能太大, 也就是说在图2的托盘中^b。作为 $S(\infty) = N - R(\infty)$ 和

$$S(T) = N - E(T) - I(T) - R(T) \simeq N - \mathcal{R}_0 R(T),$$

我们有

$$N - R(\infty) \simeq [N - \mathcal{R}_0 R(T)] \exp\left(-\frac{\mathcal{R}_0}{q N} (R(\infty) - R(T))\right).$$

作为 $R(T)/N$ 应该很小, 我们有 $\exp\left(\frac{\mathcal{R}_0 R(T)}{q N}\right) \simeq 1 + \frac{\mathcal{R}_0 R(T)}{q N}$ 和

$$1 - \frac{R(\infty)}{N} \simeq \left[1 - \frac{\mathcal{R}_0 R(T)}{N} + \frac{\mathcal{R}_0 R(T)}{q N}\right] \exp\left(-\frac{\mathcal{R}_0}{q} \frac{R(\infty)}{N}\right).$$

金牌 $\mathcal{R}_0/q < 1$ 通过假设。所以解决方案 $R(\infty)/N$ 根据经典的图形论证[5], 该方程的方程在1之前很小。使用那个 $\exp\left(-\frac{\mathcal{R}_0}{q} \frac{R(\infty)}{N}\right) \simeq 1 - \frac{\mathcal{R}_0}{q} \frac{R(\infty)}{N}$, 我们终于得到了

$$R(\infty) \simeq R(T) \mathcal{R}_0 \frac{1 - 1/q}{1 - \mathcal{R}_0/q}. \quad (14)$$

什么时候 $q \rightarrow +\infty$, 我们发现 $R(\infty) \simeq R(T) \mathcal{R}_0$ 。我们还注意到 $(1 - 1/q)/(1 - \mathcal{R}_0/q) > 1$, 应该是这样。

图3比较了 T 固定如图2^a 对于不同的参数值 q , 该流行病的最终规模以对数标度 $\log(R(\infty)/N)$ 通过公式 (14) 的数值模拟获得 $q > \mathcal{R}_0$ 并用方程的严格正解

$$1 - \frac{R(\infty)}{N} \simeq \exp\left(-\frac{\mathcal{R}_0}{q} \frac{R(\infty)}{N}\right) \quad (15)$$

为 $q < \mathcal{R}_0$ 。实际上, 在后一种情况下, 存在重大流行病, 足以适应方程 (12), 该方程或多或少与初始感染分数无关 $1 - S(T)/N \simeq \mathcal{R}_0 R(T)/N$ 只要它很小, 我们就假设 我们看到, 这两个近似在附近不再成立。 $q = \mathcal{R}_0$ 。

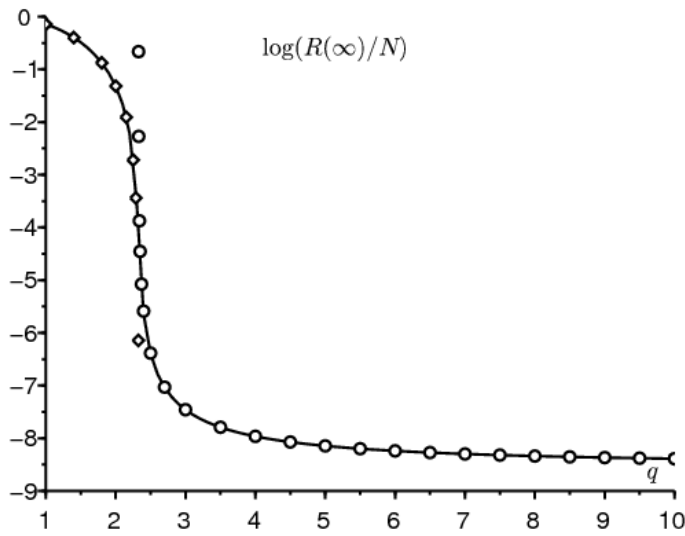


图3. $\log(R(\infty)/N)$ 取决于 q (实线), 与公式 (14) (小圆圈) 比较, 对 $q > \mathcal{R}_0$ 并使用等式 (15) 的解 (小菱形) 对 $q < \mathcal{R}_0$.

请注意, 当 q 接近 \mathcal{R}_0 . 由于很难量化 q , 因此很难预测该病的最终规模. 只有在 q 明显高于 \mathcal{R}_0 公式 (14) 的预测对 q

4. 尝试估算 q

让我们回到图1. 我们尝试估算一下 q 通过在3月15日之后将模型的仿真拟合为数据 为此, 我们从数据开始 $R(T) = 5427$. 由于前8天的数据特别对齐, 因此我们开始使用以下方法对模型进行仿真 $R(T-\tau) = e^{-\lambda\tau}R(T)$ 在哪里 $\lambda = 0,225$ 每天和 $\tau = 8$ 天, 并有相应的估算 $I(T-\tau) \simeq \frac{\lambda}{c}R(T-\tau)$, $E(T-\tau) \simeq (\frac{\lambda^2}{bc} + \frac{\lambda}{b})R(T-\tau)$ 和 $S(T-\tau) = N - E(T-\tau) - I(T-\tau) - R(T-\tau)$. 对于 $t > T$, 有效接触率为 a/q 我们尝试调整 $R(t)$ 直到3月29日为止的数据. 最合适的是 $q = 1,5$ 和 $q = 2$, 周围 $q = 1,7$. 由于此值小于 \mathcal{R}_0 , 看来遏制措施仍然不足. 也许该模型过于简单化了. 特别地, 期望在不同隔间中花费的时间的非指数分布将影响曲线开始弯曲的时刻.

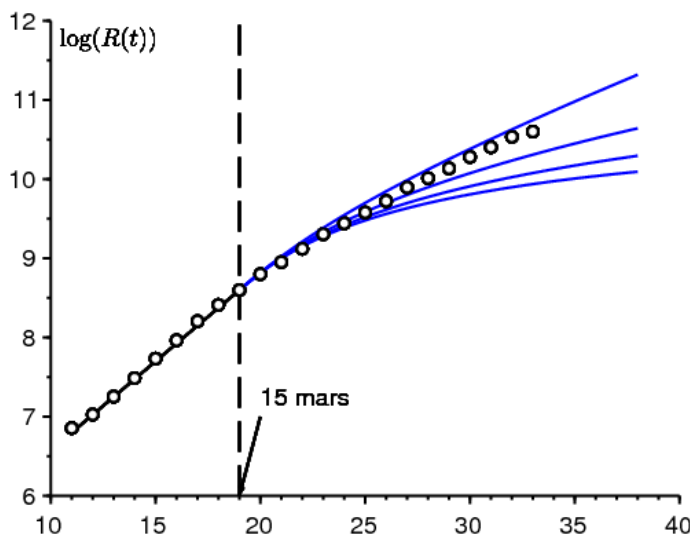


图4. 3月7日至29日 (小圆圈) 之间的病例数的自然对数 $\log(R(t))$ 作为时间的函数 t 在4个模拟中, 从上到下 $q \in \{1,5; 2; 2,5; 3\}$.

总之, 我们探索了一个两阶段方案, 其中联系率从某个日期开始降低. 我们基于减少时发现的病例数找到了该流行病最终规模的简单近似公式. 但是, 它有待更严格地陈述和证明这一结果, 可能是通过使其在渐近时表现为渐近结果来实现的. $N \rightarrow +\infty$.

致谢

我们感谢 Hisashi Inaba, Ali Moussaoui 和 Frédéric Hamelin 他们对稿件的评论。

附件

考虑具有传染周期的SIR模型，该传染周期不一定呈指数分布。要么 $I(t, x)$ 自此以来感染人口的密度 x 时间单位到时间 t 。要么 $a(x)$ 有效接触率和 $b(x)$ 感染者停止传播感染的速度。我们已经开始流行

$$\begin{aligned} I(t, 0) &\simeq \int_0^\infty a(x) I(t, x) dx \\ \frac{\partial I}{\partial t} + \frac{\partial I}{\partial x} &= -b(x) I(t, x) \\ \frac{dR}{dt} &= \int_0^\infty b(x) I(t, x) dx \end{aligned}$$

我们根据洛特卡的稳定人口理论[6]推论得出

$$I(t, x) \simeq k e^{\lambda t} e^{-\lambda x - \int_0^x b(y) dy}$$

在哪里 k 是一个常数，增长率 λ 是方程的唯一解决方案

$$1 = \int_0^\infty a(x) e^{-\lambda x - \int_0^x b(y) dy} dx.$$

如果我们问 $I(t) = \int_0^\infty I(t, x) dx$ ，问题是要估算 $I(T) + R(T)$ 来自 $R(T)$ 。金牌

$$\lambda R(T) \simeq \frac{dR}{dt}(T) = \int_0^\infty b(x) I(T, x) dx \simeq \int_0^\infty b(x) k e^{\lambda T} e^{-\lambda x - \int_0^x b(y) dy} dx.$$

我们推断

$$k \simeq \frac{\lambda R(T) e^{-\lambda T}}{\int_0^\infty b(x) e^{-\lambda x - \int_0^x b(y) dy} dx}.$$

最后，

$$\frac{I(T) + R(T)}{R(T)} \simeq \frac{\lambda \int_0^\infty e^{-\lambda x - \int_0^x b(y) dy} dx}{\int_0^\infty b(x) e^{-\lambda x - \int_0^x b(y) dy} dx} + 1.$$

我们看到右边的成员没有特别的理由与 $\mathcal{R}_0 = \int_0^\infty a(x) e^{-\int_0^x b(y) dy} dx$ 。在速率恒定的特殊情况下， $a(x) \equiv a$ 和 $b(x) \equiv b$ 但是， $\lambda = a - b$ 因此 $(I(T) + R(T))/R(T) \simeq \frac{\lambda}{b} + 1 = \frac{a}{b} = \mathcal{R}_0$ 。

参考文献

1. Bacaër N., Gomes M. G. M. (2009) Sur la taille finale des épidémies avec saisonnalité. Bull. Math. Biol. 71 : 1954 – 1966. <https://hal.archives-ouvertes.fr/hal-01299608>
2. Bacaër N. (2020) Sur le pic épidémique dans un modèle S – I – R, <https://hal.archives-ouvertes.fr/hal-02518993>
3. Corlosquet – Habart M., Janssen J., Manca R. (2012) Modélisation stochastique du risque de pandémie : stratégies de couverture et d'assurance. Lavoisier, Cachan.
4. Guan W. J. et coll. (2020) Clinical characteristics of 2019 novel coronavirus infection in China. New England Journal of Medicine, doi : 10.1056/NEJMoa2002032
5. A. Hillion (1986) Les Théories mathématiques des populations. Presses Universitaires de France, Paris.
6. Lotka A. J. (1939) Théorie analytique des associations biologiques, 2e partie. Hermann, Paris.
7. Nkague Nkamba L. (2012) Robustesse des seuils en épidémiologie et stabilité asymptotique d'un modèle à infectivité et susceptibilité différentielle. Thèse, Université de Lorraine et Université Gaston Berger.
8. Pressat R. (1995) Éléments de démographie mathématique. AIDELF, Paris.