

新型——新型冠状病毒的传播与数学的思考

复旦大学数学学院 吴宗敏

大家都在讨论新型冠状病毒的传播规律，大多数人采用的是比较宏观形象思维的方法，就是抓住几个大的影响因子事件，讨论其影响的大小深远及其叠加。这当然是属于科学的方法。而且形象思维的方法是可以精确化的，但要取决与你对重大突发事件规模的把控与掌握。因为任何事件的走向，形象地就像个树杈，你如果可以知道每个新的支叉生长点的位置与生长方向，那么，你就可以准确地描述这个树杈了。另一种就比较偏向于微观逻辑思维了，也就是看前面走过的路，来判断后面可能的走向。还是回到树杈，数学一般会引进导数。又要分几种，一种就是用前面很短的一段路程的走向来判断以后的路，这一般很难推断出以后的新分叉的时间节点在哪里？另一种是已经看见几个叉折点了，根据前面的折向规律，分析以后什么时候会产生新的叉折点，折向何处。下面，就从最简单的讲起。

刚开始时，数据很少或者人们不愿意看一大堆数据，更喜欢有意义的单个或几个数据。这样的问题，人们没有别的方法，通常采用的叫做比例原则或者说线性模型。我们从一些例子开始

1. 比例原则：

武汉封城后，某天。数据显示武汉一千四百万人，被封九百万，其中约 2000 人患病，外出约五百万人，其中约 4000 人患病。

网上就开始传，根据比例原则，这个数据造假，五百万人的得病数比九百万中的得病数的还多，比率不对，武汉的得病数肯定隐瞒了！

对吗？这个算法思想就是基于比例原则，是假设在武汉时有一个得病比例，然后，500 万人向外传，900 万人内部传，传播的比例是一样的。

这样的假设显然是错的！在外的 500 万人肯定流动性更强，或者说，传染率更大。根据当时的数据，外部传的比例与内部传的比例约为 1:4。也就是说在内部，一个病人在一定时间内传一个人的话，在外部要传四人。这显然是可信度非常高的数据。随着时间的推移，这个比例还会扩大。因为内部会变得更多的是病人传病人，而外部对病毒来说有更好的生存发展空间。

另外一个网上传言是根据日本撤侨，说日本撤侨约 200 人，其中约 10 人得病，所以武汉内部 900 万人，推断武汉约有 40~50 万人得病。这又是一个“以点带面”的典型。在武汉有与传播史密切相关的日中友好医院，日本撤的侨民中，有很大的可能，会有与病人直接接触或间接接触的人员。你不能从一个患病家庭中的患病比率推断他住的小区的患病比例，也不能用一个小区的患病比例推断整个武汉的患病比率乃至全国的患病比例，对吧？

对于大多数人那些曲线可能都不想看或反应看不懂，那就回到关注重要事件点。1. 死亡人数与治愈人数出现交叉。那表示对疾病的严重程度的描述。当然那只是表示了这个病的恶性程度，并不表示新增病例会增加还是减少。2. 传播率从小变大，而保持平稳，那可能表示是最大的传播率节点，同时表示接下来传播率会以很大的概率下降。当然可能也是传播率最大的时候。



2. 几何传播：

数据多了些，有几天的数据了，人们就可以比较前后天的数据变化。也就是我们有了一个与时间相关的函数 $f(t)$ 。根据上一节的传播率的概念，应该 $f(t+1)=c*f(t)$ 。其中 c 就称为传播率。公布数据较早就宣布了，估计 $c < 2.3$ ，比非典的 3 来的小。可是网上还是大量地在传或计算，大的时候连续地显示为 2.1 左右。可是给人们的印象是永远地这么 2.1 下去了。谣言？也不能算谣言，但定性成误导显然是合理的。

如果 c 传播率是常数，那第一个病人是从哪里来的？几何增长的速度是非常快的。网上有大量的这样的模型出现，我一般会问一句，你的模型，半年后的数据是什么？得到的病人数大大地超过了地球总人口的人数，那显然是错误的！所以，几何增长，既不能描述这个病从哪里来？也不能描述这个病会到哪里去？是一个非常粗糙的模型。

传播率一定是变化的，数学上叫做变系数的，一定是从零开始逐渐变大，达到高峰后逐渐变小，最终归零。也就是 c 应该写成 $c(t)$ ，是与时间有关的。早期公布的数字 $c(t)$ 小于 2.3 还是非常合理的，而且是非常地有预见性的。当看到 c 持续地在 2.1 那里徘徊，那是一件非常值得高兴的事情，表示它可能已经到了最高点，不大可能再会升高了。

有些模型考虑更加深刻一些，加进了克服病毒的阻力，列出了方程 $f(t)$ 的两阶导数是某常数减去 $f(t)$ 导数项。也就是病人数的的发展越大，那没受感染的人越少，从而想感染别人的机会越少，碰

到的人已经感染了，从而新增的病例会越少。这样画出了一条渐近线，说病人数达到某值时就不会再增加了。这显然时更为进步的模型。最早牛顿就考虑了这个问题。由于空气的阻力，物体的下落不是自由落体，而会有个终极速度。一个人从飞机上跳下来与从 10 层楼上跳下来碰到地面时，摔死的可能性一样。

不过这还不够！

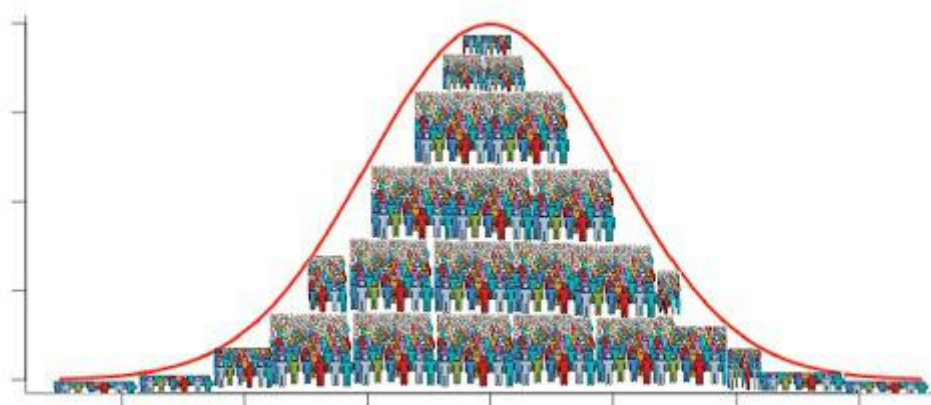
那些模型设计者说，那是牛顿的思想，已经从趋于无穷大变成趋于常数了，你还说不够？

是的！不够！得病者不可能一直保持某常数，简单地回怼是：你难道还不让那些得病者百年以后老死？得病者人数一定是一个产生，发展，减弱，消亡的过程。任何事物都是这样的生长消亡过程。不可能保持常数。

3. 高斯模型或正态模型

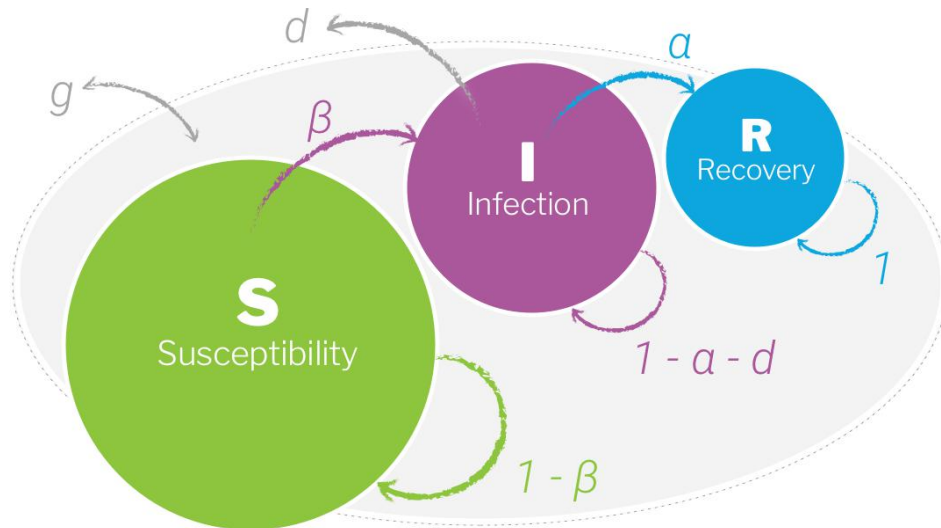
患病人数曲线一定像一座山。可能有几个山头构成，简化以后可以看成一盒山头高斯函数的形式，当然更为简单地可以写成：一个常数 A 乘以 t 的平方再乘以 $(T-t)$ 的平方，其中 T 为病疫终止点。高斯分布是对称的，即病情的发展与病情的减弱是对称的。Poisson 更为仔细地研究了这个问题，他是从顾客排队的长度入手研究这样的相似的问题。他认为患病人数应该可以写成常数 A 乘以 t 的 k 次方再乘以 $\text{Exp}(bt)$ 。也可以简化成 Beta 分布，常数 A 乘以 t 的 k 次方再乘以 $(T-t)$ 的 l 次方。也就是说，那座山可能不是对称的，可能偏向一边，歪向一边。

人们可以取不同的系数，次数来画出不同的曲线，用以拟合已知的数据，从而分析发展的可能性，找出那座山的方向，偏向度，更为重要的是那座山究竟有多大？



4. 传染病微分方程模型

也称为动力系统模型。所谓的抓住主要矛盾与主要矛盾方面的微分方程关系。根据病毒的发展力度，反病毒的措施时间力度。列出微分方程，模拟病毒的发展过程。这类方程，通常是与时俱进的，譬如，有了封城的决定，实施封城动作的时间点，间隔，都会对后续发展产生极大的影响。



5. 学习，深度学习模型

上述微分方程动力系统的方法，一般解空间或解的表达形式可以与第3节中的方法建立对应。比较困难的是：究竟那些项是主要矛盾与主要矛盾方面？学习，则是从数据出发，学习得到，哪些是关键重要的项？什么是这些项之间的关系？

6. 方法的融合，叠加与创新。

上面的方法，都是假设过程曲线是一个山头。事实上，一个新的措施，一个新的药品或者一个毒王，都会极大地改变曲线的走向，呈现多山头的现象。一般要采用多种方法的融合叠加，可不一定是线性叠加哦。在这里就有许多期待创新的想法。

附录：一些讨论

如果简单的概率估计小区有一万人，有一个人得了病，那你得病概率是不是 $1/10000$ 呢？显然不是要看这个人是谁，是家人吗，这个人是不是已经被隔离。他面对碰到你时这个人是不是打喷嚏，打喷嚏的时候是不是对着你，如果是对着你的话他自己有没有带口

罩（更重要）？你有没有戴口罩，有没有戴眼镜，有没有立即回家洗澡，用手遮挡后有没有揉眼睛。这是一个非常的不确定性的问题。

去抢购口罩，可能是一件得不偿失的行为。接触到患病者的概率会更大。

2月5日曲线的拐点好像出现了。也就是两阶导数接近零了，但也是导数或者说传播率可能最大的时候。

2020.02.06