

基于最大熵原理可靠性分析的病毒传播模型

杜亦牧^{1*}, 孙昌璞^{1,2*}

1. 中国工程物理研究院研究生院, 北京 100193;

2. 北京计算科学研究中心, 北京 100193

* 联系人, E-mail: ymdy@gascaep.ac.cn; cpsun@csrc.ac.cn

2020-04-04 收稿, 2020-04-17 修回, 2020-04-17 接受, 2020-04-20 网络版发表

国家重点研发计划(2016YFA0301201)、国家自然科学基金(11534002)和国家自然科学基金联合基金(U1930403, U1930402)资助

摘要 应用基于最大熵原理的复杂系统可靠性分析方法对病毒在宿主、易感染者和移除者等共存系统中的传播问题进行了系统的动力学理论分析. 在本研究中, 我们将病毒传播问题中的病死速率、发病速率和治愈速率等实际统计参量映射成可靠性理论中描述个体行为的风险(退化)函数: 个体宿主的病死速率和治愈速率对应于可修复系统的失效率和修复率; 潜伏期阶段患者发病速率对应于可靠性理论中的失效率. 通过最大熵原理, 本文研究融合了病死时间、治愈周期和潜伏期等统计量的各阶矩信息, 从而推断出最概然的病死速率、发病速率和治愈速率. 最后, 以包含易感、感染(确诊与潜伏期)和移除四类人群的传播模型(SEIR模型)为例, 结合基于最大熵原理的退化函数推断, 分析了新型冠状病毒肺炎(Corona Virus Disease 2019, COVID-19)传播的一些理论问题.

关键词 最大熵原理, 可靠性理论, 病毒传播模型, 新型冠状病毒肺炎

2019年末, 一例新型冠状病毒肺炎(corona virus disease 2019, COVID-19, 简称新冠肺炎)感染病例在湖北武汉被发现. 新型冠状病毒在短时间内蔓延至世界各地, 一时间成为国际社会高度关注的全球性公共卫生事件. 从理论模型的角度研究病毒传播问题具有重要的现实意义; 预测病毒传染的行为不仅是传染病和公共卫生领域的专门课题, 也是各相关领域专家学者所共同关心的问题之一.

人们对病毒传播演化问题的研究通常借助于两类模型: 确定性模型和随机性模型^[1]. 刻画病毒传播系统的演化通常采用“唯象”的非线性动力学方程, 它包含易感染人群、感染人群等动力学参量. 各类人群中的每一个人, 都直接或者间接地参与病毒传播的过程, 但从任何个体和群体角度看, 病毒传播的信息都是不完备的; 信息的缺失使得描述传播的动力学变量必须被视为随机的. 例如, 一位易感者与一位感染者从接触到

被感染的时间、被感染后到发病的时间、从发病开始到治愈或是死亡的时间、感染者所接触的易感者人数等都是不一样, 都是随机的. 所有参与病毒传播的个体构成了一个复杂系统, 动力学参数的随机性使得整个传播过程变得极为复杂, 因此随机性模型是普适的. 病毒传播确定性模型是对这类复杂系统宏观运动规律的平均化, 它假设不同个体变量的随机分布是相同的, 只能决定病毒传播的宏观行为.

根据各类人群总数随时间变化历史的信息, 通常的病毒传播分析方法是先拟合确定动力学方程中唯象参数的取值, 然后预测未来的走势(如关于新冠肺炎传播动力学的研究^[2-6]). 这类方法的优点是能够充分利用宏观统计信息给出相对准确的预测结果. 但是, 在实际应用中这类方法获得宏观信息所需要的样本数较大. 然而, 在疫情初期通常难以获得充足的历史统计数据. 最近, 有一些研究采用随机性模型, 利用统计推断确定

引用格式: 杜亦牧, 孙昌璞. 基于最大熵原理可靠性分析的病毒传播模型. 科学通报, 2020, 65: 2356-2362

Du Y M, Sun C P. Maximum entropy approach to reliability analysis based epidemic disease model (in Chinese). Chin Sci Bull, 2020, 65: 2356-2362, doi: 10.1360/TB-2020-0366

模型参数. 例如, 文献[7]利用统计推断方法研究了离散时间随机SEIR(susceptible-infective-exposed-resistance, 易感-发病-潜伏-免疫)模型. 然而, 这种方法中需要事先假设参数分布, 即事先给定似然函数, 然后才能根据已有数据推测出参数分布. 这一方法通常被应用于信息缺失的病毒传播问题中^[8], 是一种小子样不确定性分析^[9].

事实上, 由于真实病毒传播过程十分复杂, 通常的观测信息本身就是不完整的. 对于专业研究者来说, 由于技术能力和管理规制的限制, 他们面对的统计数据通常是小子样事件样本. 因此, 即便是采用确定模型, 也应该适当考虑信息不完整带来的不确定性. 要处理这一问题, 必须考虑从“唯象”到“微观”这一关键问题: 在求解传播方程之前, 利用统计方法确定刻画基本传播过程的变量分布.

最近, 对于复杂物理系统的可靠性研究, 我们建立了基于最大熵原理的可靠性理论^[10], 以描述个体行为的风险(退化)函数. 现在, 我们把病毒传播问题中的病死速率、发病速率和治愈速率等实际统计参量与风险(退化)函数对应起来, 从而把基于最大熵原理的可靠性分析理论应用到病毒传播相关变量分布的确定. 与文献[2~6]中的方法不同之处在于, 我们的方法更加侧重于利用个体统计信息而非总体人数随时间变化的信息, 并且通过统计推断方法把信息不完整这一要素融合到了模型当中. 这一不同处理会带来以下潜在优势: (1) 在一定程度上减少所需样本的数量; (2) 使得传播动力学方程中的一部分唯象参数不再需要通过拟合方式获得, 增加拟合的效率; (3) 从微观角度得到的变量随机分布不依赖传播模型, 这些结果适用于多种模型. 当然, 我们的方法也存在一些缺点, 例如小子样统计会损失一定的精确度.

作为有现实意义的应用, 本文采用文献[11,12]中的统计数据, 通过拓展SEIR模型分析了新冠肺炎的传播行为. 在考虑有限免疫时间的情况下, 本文的研究给出了最终稳态感染人数比例与免疫治愈时长比之间的关系, 在能够获取免疫时长信息的前提下, 可以用作预测系统平稳之后携带病毒者的比例.

1 最大熵可靠性分析确定病死速率、治愈速率及发病速率

在基于最大熵原理的可靠性分析中^[10], 我们利用平均寿命及其涨落等各阶矩信息确定系统的最概然失

效率, 使得人们可以从统计物理基本理论的角度理解各类失效率曲线. 其实, 这个理论可以用来确定病毒传播模型中的关键参数及其随机行为.

先考虑感染者的治愈率和病死率. 把感染者被确诊到被治愈的时间记为 t , 被治愈后的余留寿命记为 t' . 由于个体患者之间的差异, 这两个特征量可视为随机变量, 且满足待定分布:

$$\begin{aligned} p(t, t') &= (1-\mu)p_{RD}(t, t') + \mu p_D(t)\delta(t') \\ &\equiv (1-\mu)p_R(t)p_{D|R}(t, t') + \mu p_D(t)\delta(t'), \end{aligned} \quad (1)$$

其中, p_{RD} 是治愈时间和余留寿命的联合概率分布, 满足 $p_{RD}(t, t') = p_R(t)p_{D|R}(t, t')$, 即经历了 t 时间的治疗并且在治疗之后 t' 时间死亡的概率, 满足 $\int\int p_{RD} dt dt' = 1$; $p_{D|R}(t, t')$ 是相应的条件概率, 即治愈时间为 t 的条件下治愈后 t' 时间死亡的概率, 满足 $\int p_{D|R}(t, t') dt' = 1$; p_R 是治愈时间的概率分布, 满足 $\int p_R(t) dt = 1$; p_D 是病死时间的概率分布, 满足 $\int p_D(t) dt = 1$; δ 函数 $\delta(t')$ 意味着患者病死; μ 是病死人数占总感染人数的比例. 根据可靠性理论, 各种特征时间的分布可以通过失效率函数来表示: $p_\alpha = x_\alpha(t)\exp(-Q_\alpha)$, 其中, $Q_\alpha = \int_0^t x_\alpha(t') dt'$; $x_\alpha(t)$ 是失效率函数, 指标 α 取R, D分别表示治愈速率和病死速率.

本工作的关键是通过最大熵原理给出这些分布的具体函数形式. 考虑到病死与治愈两种事件可以分辨, 分布(1)的信息熵是

$$\begin{aligned} S_{RD} &= -(1-\mu) \int_0^{+\infty} \int_0^{+\infty} p_{RD}(t, t') \ln p_{RD}(t, t') dt' dt \\ &\quad - \mu \int_0^{+\infty} p_D(t) \ln p_D(t) dt - (1-\mu) \ln(1-\mu) - \mu \ln \mu. \end{aligned} \quad (2)$$

把治愈时间和病死时间的各阶矩作为约束条件

$$\begin{aligned} C_{RD} &= -(1-\mu) \sum_i \beta_{R,i} \int_0^{+\infty} p_R(t) t^i dt \\ &\quad - \mu \sum_i \beta_{D,i} \int_0^{+\infty} p_D(t) t^i dt, \end{aligned} \quad (3)$$

其中, $\beta_{R(D),i}$ 是随机变量 t 第 i 阶矩约束的拉格朗日乘子. 这里的约束条件与 p_{RD} 无关是因为在接下来的讨论中并未考虑余留寿命的信息.

最大熵原理要求分布函数满足约束条件且使熵取值最大, 即 $\delta(S_{RD} + C_{RD}) / \delta Q_{R(D)} = 0$. 求解欧拉拉格朗日方程后, 得到治愈速率和病死速率如下:

$$x_\alpha = \frac{x_\alpha(0)\exp\left(-\sum_i \beta_{\alpha,i} t^i\right)}{1 - \int_0^t x_\alpha(0)\exp\left(-\sum_i \beta_{\alpha,i} \tilde{t}^i\right) d\tilde{t}}, \quad (4)$$

其中, 指标 α 取 R, D, $x_\alpha(0)$ 既是初始时刻治愈 ($\alpha=R$)、病死 ($\alpha=D$) 速率, 也是分布的归一化系数. 通过这种方式得到的病死速率、治愈速率和相应的概率分布只与观测数据有关, 与具体的病毒传播模型无关. 类似地, 考虑易感染者(未感染者)接触传染源被感染到发病的时间 t (潜伏期时长), 将 t 看成随机变量, 满足分布 $p_1 = x_1(t)\exp(-Q_1)$, 其中 x_1 表示发病速率, 且 $Q_1(t) = \int_0^t x_1(t') dt'$. 在各阶矩约束下最大化分布 p_1 的信息熵, 即使得如下形式的泛函 \tilde{S}_1 取极值:

$$\tilde{S}_1 = S_1 + C_1 = - \int_0^{+\infty} p_1 \ln p_1 dt - \sum_i \beta_{1,i} \int_0^{+\infty} p_1 t^i dt, \quad (5)$$

其中, $-\int_0^{+\infty} p_1 \ln p_1 dt$ 是分布 p_1 的信息熵; $\beta_{1,i}$ 是随机变量 t 的第 i 阶矩约束的拉格朗日乘子. 通过求解 $\delta(S_1 + C_1) / \delta Q_1 = 0$, 可以计算得到发病速率

$$x_1 = \frac{x_1(0)\exp\left(-\sum_i \beta_{1,i} t^i\right)}{1 - \int_0^t x_1(0)\exp\left(-\sum_i \beta_{1,i} \tilde{t}^i\right) d\tilde{t}}, \quad (6)$$

其中, $x_1(0)$ 既是初始时刻治愈速率, 也是归一化常数, 满足 $\int_0^{+\infty} x_1(0)\exp\left(-\sum_i \beta_{1,i} t^i\right) dt = 1$. 同样地, 通过这种方式得到的发病速率只与观测数据有关, 与具体传播模型无关. 通过最大熵原理得到了3种影响病毒传播的速率, 可以将这些结果应用于传播模型中.

2 病毒传播模型

在上一节中, 我们阐述了如何利用最大熵原理求得最概然的病死速率、治愈速率及发病速率. 这些退化函数能够直接应用于病毒传播的研究. 我们将从以下3个假设出发建立病毒传播模型: (1) 有四类人群: 易感染者、确诊患者、处于潜伏期的患者与康复者, 这些人群的人数依次记为 N_H, N_V, N_{VH}, N_R , 并且假设康复

者具有免疫能力. (2) 与易感染者接触的人携带病毒概率等于病毒携带者(确诊患者或处于潜伏期的患者)人数占总人数的比例 $(N_V + N_{VH}) / (N_H + N_V + N_{VH} + N_R)$. 每个易感染者每天接触的平均人数为 n_0 , 假设 n_0 不依赖于人口总数, 且受人为干预措施的影响. 这里, n_0 是个唯象参数. (3) 易感染者接触病毒携带者后被感染的概率为 v , 且每次接触后感染的概率相互独立.

基于以上三个假设, 病毒传播模型可以由以下的一组微分-积分方程组表达:

$$\begin{cases} \dot{N}_H = -\phi_0, & (7.1) \\ \dot{N}_V = \phi_1 - \phi_2 - \phi_3, & (7.2) \\ \dot{N}_{VH} = \phi_0 - \phi_1, & (7.3) \\ \dot{N}_R = \phi_2, & (7.4) \end{cases} \quad (7)$$

其中, 方程(7.1)表示易感人数减少速率, 它等于单位时间内被感染的人数(ϕ_0). 它可以通过如下方式得到: 考虑一个易感者在单位时间内接触人数 n_0 , 每个接触者携带病毒的概率是病毒携带者人数占总人数的比例; 一旦易感者接触到病毒携带者, 被感染的概率是 v . 于是, 单位时间内 N_H 个易感者中平均将有 ϕ_0 个人被感染:

$$\begin{aligned} \phi_0 &= \lim_{\tau \rightarrow 0} \frac{N_H}{\tau} \left(1 - \left[1 - \frac{v(N_V + N_{VH})}{N_H + N_V + N_{VH} + N_R} \right]^{n_0 \tau} \right) \\ &\approx \frac{n_0 v N_H (N_V + N_{VH})}{N_H + N_V + N_{VH} + N_R}, \end{aligned}$$

其中, τ 是无穷小时间间隔, $n_0 v \approx R_0 / T_S$, R_0 是每个感染者平均传染人数, T_S 是人际传染间隔(或称世代间隔), 近似中考虑了 $v \ll 1$ 的条件. 方程(7.2)表示发病(确诊)人数随时间的变化, 它等于单位时间内新增发病(确诊)人数(ϕ_1)减去单位时间内治愈(ϕ_2)和病死(ϕ_3)的人数,

$$\phi_1 = \int_0^t p_1(t-t') \phi_0(t') dt' + N_{VH}(0) p_1(t),$$

$$\phi_2 = \int_0^t (1-\mu) p_R(t-t') [\phi_1(t') + N_V(0) \delta(t')] dt',$$

$$\phi_3 = \int_0^t \mu p_D(t-t') [\phi_1(t') + N_V(0) \delta(t')] dt'. \quad \text{方程(7.3)表示处在潜伏期的人数的时间变化率, 它等于单位时间内被感染的人数}(\phi_0)$$

减去单位时间内发病的人数(ϕ_1). 方程(7.4)表示治愈人数随时间的增长速率.

3 通过失效率分析病毒传染行为

第2节考虑了带时滞的SEIR(易感-发病-潜伏-免疫)

模型,且模型中所需的概率分布正是通过最大熵原理给出的概率分布.类似地,如果考虑其他传播模型,例如SIS(易感-感染-易感)、SIR(易感-发病-免疫)等,这些概率分布依然适用.从上述模型出发,本节分析感染初期感染人数增长速率与病死速率、治愈速率及发病速率之间的依赖关系以及方程的渐进行为.假设在观测中得到的信息只有一阶矩,即最大熵原理给出的分布是一阶矩模型.一阶矩模型仅考虑平均病死时间 T_D 、平均治愈周期 T_R 和平均潜伏期长度 T_I .此时,如果不加额外信息约束,由方程(4)与(6)最大熵原理将给出分布函数

$$p_\alpha = \exp(-t/T_\alpha)/T_\alpha, \quad (8)$$

以及相应的失效率 $x_\alpha = 1/T_\alpha$,其中, $\alpha = R, D, I$.

由概率分布(8),得到 ϕ_1, ϕ_2 和 ϕ_3 对时间的变化率

$$\begin{cases} \dot{\phi}_1 = \phi_0/T_I - \phi_1/T_I, & (9.1) \\ \dot{\phi}_2 = (1-\mu)\phi_1/T_R - \phi_2/T_R, & (9.2) \\ \dot{\phi}_3 = \mu\phi_1/T_D - \phi_3/T_D, & (9.3) \end{cases} \quad (9)$$

其中, ϕ_1, ϕ_2, ϕ_3 分别表示单位时间新增确诊人数

$$\phi_1 = \int_0^t \exp[-(t-t')/T_I] \phi_0(t') dt' / T_I + N_{VH}(0) \exp[-t/T_I] / T_I,$$

单位时间新增治愈人数

$$\phi_2 = \int_0^t (1-\mu) \exp[-(t-t')/T_R] \phi_1(t') dt' / T_R$$

$$+ (1-\mu) \exp[-t/T_R] N_V(0) / T_R,$$

和单位时间新增死亡人数

$$\phi_3 = \mu \int_0^t \exp[-(t-t')/T_D] \phi_1(t') dt' / T_D$$

$$+ \mu \exp[-t/T_D] N_V(0) / T_D.$$

基于上述的传播方程,我们研究在传播过程中感染人数峰值相关的问题以及初期感染人数增长速率.

3.1 传播进程中感染人数的峰值与定性分析

通过以下分析可以说明当 $n_0 v$ 与时间无关时,在上述SEIR模型所描述的传播进程中感染人数的极大值只有一个.假设初始时刻感染人数较少,忽略初始感染人数 $N_{VH}(0), N_V(0)$ 的贡献,总感染人数在 $\phi_0 = \phi_2 + \phi_3$ 时达到极大值,此时

$$\begin{aligned} \frac{d^2(N_V + N_{VH})}{dt^2} &= -\frac{(N_V + N_{VH})\phi_2}{N_V + N_{VH} + N_H + N_R} \\ &\quad - \frac{(N_V + N_{VH})(N_V + N_{VH} + N_R)\phi_3}{(N_V + N_{VH} + N_H + N_R)^2} \\ &\quad - \phi_1 \left(\frac{1-\mu}{T_R} + \frac{\mu}{T_D} \right) - \frac{\phi_2}{T_R} - \frac{\phi_3}{T_D} \leq 0. \end{aligned} \quad (10)$$

这一计算结果表明,如果感染人数出现极值点,那么一定是一个极大值点.由于方程(7)中感染人数对时间是可导的,因此感染人数随时间变化只会出现唯一的一个极大值,之后随着时间增长逐步下降并趋于常数,通过对不等式(10)取等号可以看出这一常数是零.

由于SEIR模型中假定了治愈者永久免疫,传播足够长时间后系统内绝大部分都是免疫者,感染者的比例是足够低的.这表示当免疫状态持续时间远远大于治愈时间时,最终系统中感染者的比例将趋近于零.免疫时长是一个有限值的情形将在后面的数值结果中展示.这一定性结论的现实意义在于,延长免疫期对降低感染比例起到关键作用.

3.2 初期感染人数增长速率

考虑初期感染人数的增长速率.在疫情初期,治愈人数与感染人数都远远小于未感染人数,即 $N_V + N_{VH} \ll N_H, N_R \ll N_H$.此时,非线性项的主要贡献由线性部分给出:

$$\frac{(N_V + N_{VH})N_H}{N_H + N_V + N_{VH} + N_R} \approx N_V + N_{VH}.$$

把这一条件代入方程(7),并考虑到方程(8),对方程(7)作拉普拉斯变换,即方程(7)两边同乘以 $\exp(-st)$,并对 t 作0到 $+\infty$ 的积分.通过求解拉普拉斯变换得到线性方程组,可知感染人数的增长速率等于以下方程的正根:

$$(s + 1/T_I)(s + 1/T_D)(s + 1/T_R) - \gamma(s) = 0, \quad (11)$$

其中, s 是拉普拉斯变换的宗量;

$$\gamma(s) = n_0 v [s^2 + \mu x_I x_R + s(x_D + x_R + x_I) + x_D(x_I - \mu x_I + x_R)].$$

具体计算过程详见附录1.

利用方程(11)的解可以确定初期感染人数的增长速率.方程(11)有3个根,从小到大依次记为 $s_1, s_2, s_3 (s_1 < s_2 < s_3)$,利用反演公式可知 $N_V + N_{VH} = a_0 + \sum_{i=1}^3 a_i \exp(s_i t)$,其中系数 $a_i (i = 0, 1, 2, 3)$ 取决于初始条件 $N_V(0), N_{VH}(0)$ 和 $T_R, T_D, T_I, n_0 v$,与时间 t 无关.这一结果表明,在一阶矩模型中,感染人数初始增长速率近似等于 s_3 .它是通过包含各个特征时间的一阶矩、病死

比例和传播系数 n_0v 的高次方程确定的. 方程(11)中的各个参数除了 n_0v 是唯象的, 其余的都能直接通过最大熵原理得到.

4 数值结果

4.1 永久免疫的情形

在上一节的讨论中, 我们假设治愈者是永久免疫的. 对于一阶矩模型, 考虑没有人为干预的情形. 考虑目前关于新冠肺炎的部分统计数据 $T_R = 12.8^{[11]}$, $T_I = 5.2$, $T_S = 7.5^{[12]}$, 并且估计 $T_D \approx 15$, $\mu \approx 0.04$, $n_0v \approx R_0 / T_S \approx 3 / 7.5$. 对于一阶矩模型, 考虑有人为干预的情形. 考虑人为干预介入时间 t_0 以及措施平均生效时间 T , 假定 $n_0v \approx R_0 / T_S \exp(-t/T)\theta(t-t_0)$.

数值结果如图1所示. 可以看出, 感染人数随着时间不断增长达到峰值, 继而回落逐步趋近于零. 如果在上升阶段加以干预减小 n_0v , 那么峰值会提前到来, 并且峰值比例远低于无干预措施时的数值.

4.2 有限免疫时间的情形: 稳态感染比例与免疫时长的关系

考虑个体只在治愈后的一段时间内(记为 t)具有免疫能力. 假设 t 是一个随机变量, 满足分布 p_{IM} , 并且平均免疫时间为 T_{IM} . 利用这一平均值信息, 通过最大熵原理推出 $p_{IM} = \exp(-t/T_{IM})/T_{IM}$. 考虑这一条件, 将传播方程(7)改写为

$$\begin{cases} \dot{N}_H = -\phi_0 + \phi_4, & (12.1) \\ \dot{N}_V = \phi_1 - \phi_2 - \phi_3, & (12.2) \\ \dot{N}_{VH} = \phi_0 - \phi_1, & (12.3) \\ \dot{N}_R = \phi_2 - \phi_4, & (12.4) \end{cases} \quad (12)$$

其中, T_{IM} 代表免疫能力消失的平均时间. 与方程(7)不同之处在于, 方程(12)中多了与 ϕ_4 有关的项, 它的含义是单位时间内失去免疫能力的人数, 满足 $\dot{\phi}_4 = \phi_2 / T_{IM} - \phi_4 / T_{IM}$; 方程(12)中其余变量满足的条件与方程(9)中的一致. 我们计算时间趋于无穷时的稳态特征(这里时间趋于无穷指的是时间尺度远大于 T_R, T_{IM}, T_D, T_I 等基本过程的时间尺度), 结果如图2所示, 当免疫时间远大于治愈时间时, 稳态的感染者比例趋于零. 当平均治愈时间与平均免疫时间的比例上升时, 稳态的感染者比例也随之逐步上升. 实际上, 对于平均

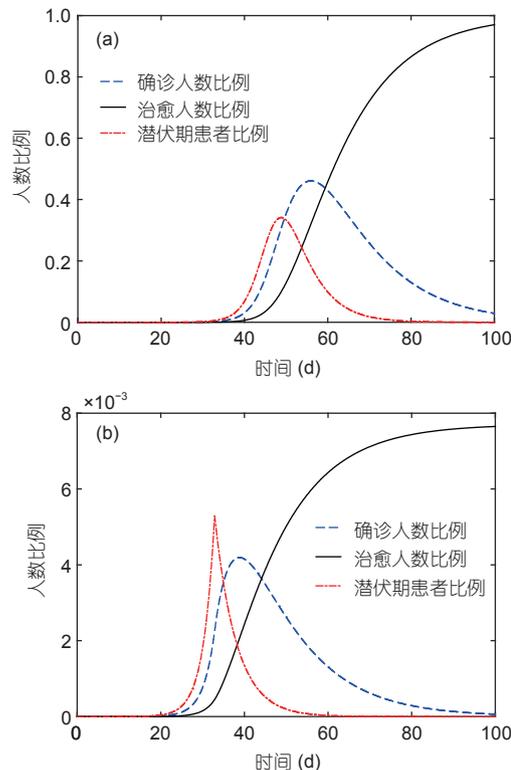


图1 (网络版彩色)确诊、潜伏期和治愈三类人数比例随时间的变化. 其中初始条件设置为 $N_H(0)=3 \times 10^7$, $N_V(0)=1$, $N_{VH}(0)=0$, $N_R(0)=0$; $t_0=33$, $T=1$. (a) 无干预措施的结果; (b) 有干预措施的结果
Figure 1 (Color online) Population ratios of diagnosed, incubation and cured vary with time. The initial conditions are set to be $N_H(0)=3 \times 10^7$, $N_V(0)=1$, $N_{VH}(0)=0$, $N_R(0)=0$; $t_0=33$, $T=1$. (a) The result without intervention; (b) the result with intervention

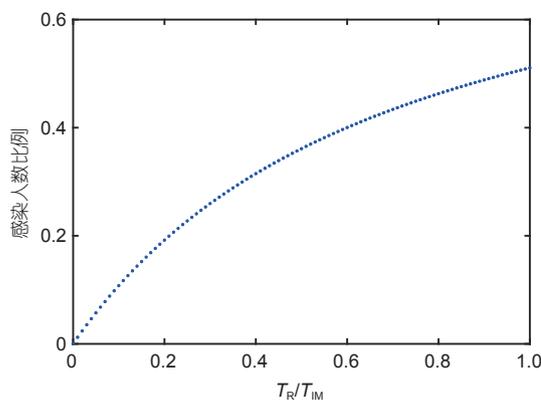


图2 (网络版彩色)一阶矩模型稳态感染比例与免疫治愈时间比的关系. 其中 T_R 和 T_{IM} 分别表示平均治愈时间和平均免疫时间
Figure 2 (Color online) The relationship between infectious ratio and the proportion of immunization duration to healing period. T_R and T_{IM} denote the average healing period and the average immunization duration, respectively

免疫时间远大于平均治愈时间的情形, 方程(12)回到SEIR模型; 对于平均免疫时间远小于平均治愈时间的

情形, 方程(12)等价于SEIS模型.

5 总结与讨论

我们研究了描述病毒传染的时滞动力学模型, 并利用基于最大熵原理的可靠性分析方法将模型中的概率分布中的病死速率、发病速率、治愈速率等唯象参数与对少量个体病例的统计量相联系. 以一阶矩模型为例, 从解析的角度讨论了方程的渐进行为和初始感染人数增长率与病死速率、发病速率、治愈速率的依赖关系. 利用已有的一些统计数据^[11,12]并通过数值计算, 展示了有人为干预措施和无人干预措施两种情形下, 发病人数、潜伏期人数、治愈人数等随时间的变化. 此外, 还讨论了免疫时间有限的情形, 给出了系统稳态感染人数比例与免疫治愈时间比之间的关系.

相较于传统方法, 该方法的优点在于侧重于利用

个体细节的统计信息而非总体人数随时间变化的信息. 该方法的特点在于: (1) 确定参数的过程不依赖于动力学模型的选取, 这是因为病死速率、发病速率、治愈速率等参数描述的是复杂系统的微观特征; (2) 只需要少量样本即可给出关于趋势、感染人数拐点等关键信息的预测, 对疫情初期的快速预测会起到重要帮助作用; (3) 可以融合各类时间间隔的不同信息, 例如各阶矩和中位数信息(附录2).

最后需要指出, 由此方法给出的预测结果的好坏会受到以下因素的影响: (1) 最大熵原理约束条件中融合各阶矩信息的数目会影响最终预测的准确程度以及细节的丰富程度; (2) 动力学模型本身也会影响准确程度. 今后我们将优化该方法, 进一步考虑结合微观细节信息和宏观人数信息等问题, 对实际问题给出更加准确的预测.

致谢 感谢与中国科学院武汉物理与数学研究所蔡庆宇教授、中国工程物理研究院研究生院董辉教授、北京计算科学研究中心马宇翰博士等人的讨论.

参考文献

- 1 Bailey N T J. The Mathematical Theory of Infectious Disease. 2nd ed. New York: Hafner, 1975
- 2 Tang B, Wang X, Li Q, et al. Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. *J Clin Med*, 2020, 9: 462
- 3 Tang B, Bragazzi N L, Li Q, et al. An updated estimation of the risk of transmission of the novel coronavirus (2019-nCoV). *Infect Dis Model*, 2020, 5: 248–255
- 4 Yan Y, Chen Y, Liu K J, et al. Modeling and prediction for the trend of outbreak of NCP based on a time-delay dynamic system (in Chinese). *Sci China Math*, 2020, 50: 1–8 [严阅, 陈瑜, 刘可伋, 等. 基于一类时滞动力学系统对新型冠状病毒肺炎疫情的建模和预测. *中国科学: 数学*, 2020, 50: 1–8]
- 5 Chen Y, Cheng J, Jiang Y, et al. A time delay dynamic system with external source for the local outbreak of 2019-nCoV. *Appl Anal*, 2020, doi: 10.1080/00036811.2020.1732357
- 6 Lin Q, Zhao S, Gao D, et al. A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. *Int J Infect Dis*, 2020, 93: 211–216
- 7 Lekone P E, Finkenstädt B F. Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. *Biometrics*, 2006, 62: 1170–1177
- 8 Hotta L K. Bayesian melding estimation of a stochastic SEIR model. *Math Popul Stud*, 2010, 17: 101–111
- 9 Zacks S. Introduction to Reliability Analysis Probability Models and Statistical Methods. New York: Springer-Verlag, 1992
- 10 Du Y M, Ma Y H, Wei Y F, et al. Maximum entropy approach to reliability. *Phys Rev E*, 2020, 101: 012106
- 11 Guan W J, Ni Z Y, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med*, 2020, doi: 10.1056/NEJMoa2002032
- 12 Li Q, Guan X, Wu P, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med*, 2020, 382: 1199–1207

补充材料

附录1 推导方程(11)的计算细节

附录2 基于中位数信息的统计分析

本文以上补充材料见网络版csb.scichina.com. 补充材料为作者提供的原始数据, 作者对其学术质量和内容负责.

Summary for “基于最大熵原理可靠性分析的病毒传播模型”

Maximum entropy approach to reliability analysis based epidemic disease model

Yi-Mu Du^{1*} & Chang-Pu Sun^{1,2*}

¹ Graduate School of China Academy of Engineering Physics, Beijing 100193, China;

² Beijing Computational Science Research Center, Beijing 100193, China

* Corresponding authors, E-mail: ymdu@gscaep.ac.cn; cpsun@csrc.ac.cn

This study analyzes the dynamic model of epidemic disease system based on the maximum entropy approach to reliability. The epidemic disease system is assumed to be composed of several classes, namely susceptible class, infective class, resistance class, exposed class and so on. The dynamics of the populations in each class are depicted by a number of differential equations, where the populations transmitting between the classes are the averages of transmitting populations in microscopic stochastic transmission process. In microscopic transmission process, several basic variables are assumed to be random, such as incubation period, hospitalization duration and so on. To infer the probability distributions of these variables is a main task of this study. Inspired by the recent study of reliability, the maximum entropy based approach is applied to determine the parameters in the dynamic model. The basic idea of this study is presented below.

In this study, degrade (hazard) function, which is a fundamental quantity in the disciplines of risk and reliability analysis, is associated with death rate, incidence rate and healing rate. Specifically, death rate and healing rate of the infective people are associated with hazard rate and repair rate of the repaired system in reliability theory, respectively; the incidence rate is associated to the hazard rate during incubation period. By means of the maximum entropy principle, the moments of the period from onset to death, the period from onset to recovery and the incubation period are fused to infer the most probable death rate, incidence rate and healing rate. Applying the maximum-entropy based statistical inference, the information of macroscopic transmission process is fused, such as average values, fluctuations and median values. To the best of our knowledge, the traditional fitting approaches to determine parameters usually rely on the information of macroscopic phenomenon. It is different from these fitting approaches that the maximum-entropy based approach applied in this study relies on the information of microscopic process. Thus this approach is adapted to practice scenario where the limitations of information access and number of samples both exist. And the parameter determination is independent on the choice of the macroscopic epidemic disease dynamic model.

After applying the maximum entropy principle based inference and the SEIR (susceptible-infective-exposed-removed) model, several discussions associated with coronavirus disease 2019 (COVID-19) are made. With the help of recent information of microscopic transmission process, the parameters in SEIR model are determined directly by the maximum-entropy principle. Then with numerical calculation, dynamics of the populations in infective class, resistance class and exposed class are obtained. Besides, some typical phenomena are revealed by the analytical and numerical results. For example, the calculation shows that the peak of the infectious ratio in transmission process is unique; the infectious ratio of steady state tends to zero when the immunity duration is much larger than the healing period. Additionally, limited immunity duration model is also considered. The relationship between the infectious ratio of steady state and the proportion of immunization duration to healing period is presented.

maximum-entropy principle, reliability theory, epidemic disease model, novel coronavirus pneumonia

doi: 10.1360/TB-2020-0366