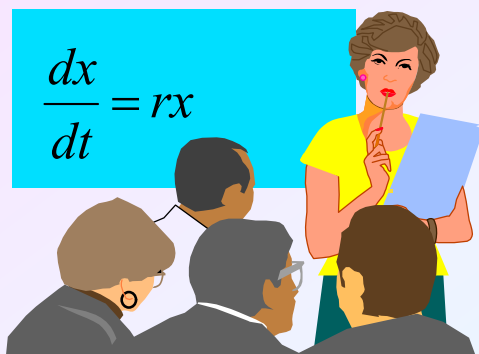


搭建数学与现实之间的桥梁

开启思维与灵感之门的钥匙

——浅谈数学建模的功能和魅力



清华大学 姜启源
jiangqy@tsinghua.edu.cn

1. 引言——**抗击疫情**中的数学建模

2. 数学建模在**科技发展和****社会进步**中的重要地位

3. 数学建模引入大学教育促进了学生**综合素质**
的培养与提高

4. 怎样学习数学建模——**学好他人的与** **做好自**
己的；对两种常用建模方法的一些看法

抗击疫情中的数学建模

核酸检测的混检方案

背景

- 2020.5.14~6.1武汉对9,899,828人进行了核酸检测，检出无症状感染者300名，阳性率为0.303/万。
- 采取25%混检+75%单检方案，混检是把5份样本混在一起检测，如果混合样本呈阳性再进行单检，既保证准确性也提高了效率（减少检测次数）。
- 检测前做了小规模测试，用于确定混检方案（将多少份样本混合检验）。

问题

- 怎样根据事先测试的结果确定混检方案？
- 混检一定能够提高效率吗？

核酸检测的混检方案

问题分析

- 若混合样本呈阳性需再单检, 检测次数增加; 若混合样本呈阴性无需单检, 检测次数减少.
- 根据事先测试设定阳性率, 建立混检平均检测次数模型, 按照平均检测次数最少确定混检方案.

模型假设

- 单个样本阳性率为 p , p 很小, 且每个样本是否呈阳性相互独立.
- 当混合样本呈阴性时, 其单个样本均不检测. 当混合样本呈阳性时, 其单个样本均做检测.
- 总样本数为 n , 混合样本数为 k (将 k 份样本混合), $n \gg k$, 共有 n/k 个混合样本 (视为整数).

核酸检测的混检方案

模型建立

单个样本阳性率 p , 阴性率 $q=1-p$, k 份样本混合.

混合样本呈阴性: 概率 q^k , 只检测 1 次;

混合样本呈阳性: 概率 $1-q^k$, 需检测 $k+1$ 次.

- 1 个混合样本的平均检测次数:

$$m_1 = q^k + (k+1)(1-q^k) = k+1 - kq^k$$

- 全部样本的平均总检测次数:

$$m = m_1(n/k) = [1 + 1/k - (1-p)^k]n \quad \sim \text{混检模型}$$

问题: 给定总样本数 n 及阳性率 p , 确定混检方案 k ,
使平均总检测次数 m 最少.

核酸检测的混检方案

$$m = [1 + 1/k - (1-p)^k]n$$

模型
求解

模型简化: p 很小 $\Rightarrow (1-p)^k \approx (1-kp)$

$$\Rightarrow m/n \approx 1 + 1/k - (1-kp) = 1/k + kp$$

\Rightarrow 当 $k \approx p^{-1/2}$ (取邻近整数) 时 m/n 最少.

p	5%	2%	1%	0.1%	0.01%
k	5	8	10	32	100
m/n	0.426	0.274	0.196	0.063	0.020

- 阳性率 p 越小, 混合样本数 k 越大, 平均检测次数越少.
- 检测前小规模测试用于估计阳性先验概率 p .
- 武汉实际阳性率 0.303/万, 即阳性后验概率 0.003%.

核酸检测的混检方案

$$m = [1 + 1/k - (1-p)^k]n$$

模型 讨论

- 混检一定能够提高效率吗？

阳性率 p 多大, 就不应采用混检方案
(平均总检测次数 m 不小于 n) .

$$m/n = 1 + 1/k - (1-p)^k \geq 1 \quad \Leftrightarrow \quad p \geq 1 - k^{-1/k}$$

- 检查 $k=2, 3$ 可得, 当 $p > 0.307$ 时不应采用混检方案.
- 当混检方案 k 较大时, 是否可作第2次混检?
建立两次混检的数学模型.
- 若模型假设“混合样本呈阳性单个样本均做检测”
改为“单个样本顺序检测”, 建模有何改变?

新冠病毒流行病趋势预测

Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions

Zifeng Yang, Zhiqi Zeng, Ke Wang, Sook-San Wong, Wenhua Liang, Mark Zanin, Peng Liu, Xudong Cao, Zhongqiang Gao, Zhitong Mai, Jingyi Liang, Xiaoqing Liu, Shiyue Li, Yimin Li, Feng Ye, Weijie Guan, Yifan Yang, Fei Li, Shengmei Luo, Yuqi Xie, Bin Liu, Zhoulang Wang, Shaobo Zhang, Yaonan Wang, Nanshan Zhong, Jianxing He

Journal of Thoracic Disease, Vol 12, No 3 March 2020

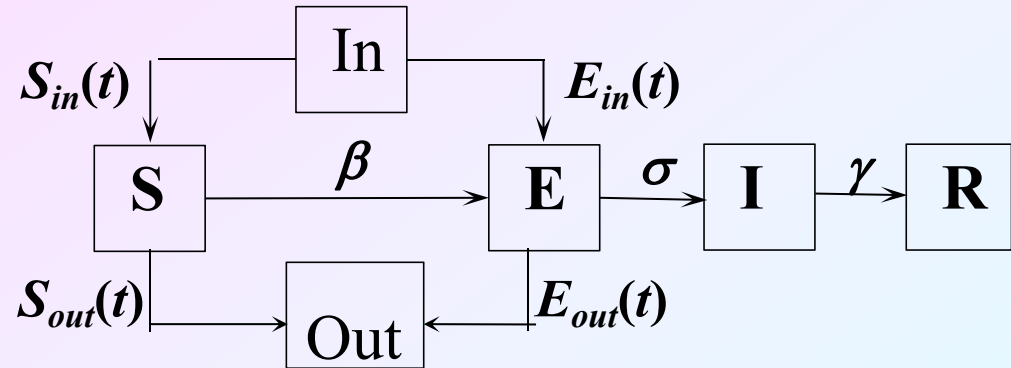
<http://dx.doi.org/10.21037/jtd>. 2020.02.64

Submitted Feb 27, 2020. Accepted for publication Feb 28, 2020

Modified SEIR model

$$\begin{aligned}\frac{dS}{dt} &= -\frac{\beta SI}{N} \\ \frac{dE}{dt} &= \frac{\beta SI}{N} - \sigma E \\ \frac{dI}{dt} &= \sigma E - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}$$

$S(t)$ ~susceptible, $E(t)$ ~exposed,
 $I(t)$ ~infected, $R(t)$ ~recovery or death



$$S(t+1) = S(t) - [\beta_1 I(t) + \beta_2 E(t)]r(t)S(t) / N(t) + S_{in}(t) - S_{out}(t)$$

$$E(t+1) = E(t) + [\beta_1 I(t) + \beta_2 E(t)]r(t)S(t) / N(t) - \sigma E(t) + E_{in}(t) - E_{out}(t)$$

$$I(t+1) = I(t) + \sigma E(t) - \gamma I(t)$$

$$R(t+1) = R(t) + \gamma I(t)$$

$r(t)$ ~ number of contacts per person per day.
Before Jan 23, $r = 15$, after Jan 23, $r = 3$,
and after March 1, $r = 10$.

β_1, β_2 ~rate of transmission for the susceptible to infected and exposed.

σ ~ incubation rate, γ ~ average rate of recovery or death.

$$E_{in}(t) = In(t)P_{in}(t), \quad S_{in}(t) = In(t)[1 - P_{in}(t)]$$

$$E_{out}(t) = Out(t)P_{out}(t), \quad S_{out}(t) = Out(t)[1 - P_{out}(t)]$$

$P_{in/out}(t)$ ~probability of the inflowing/outflowing exposed

Data sources

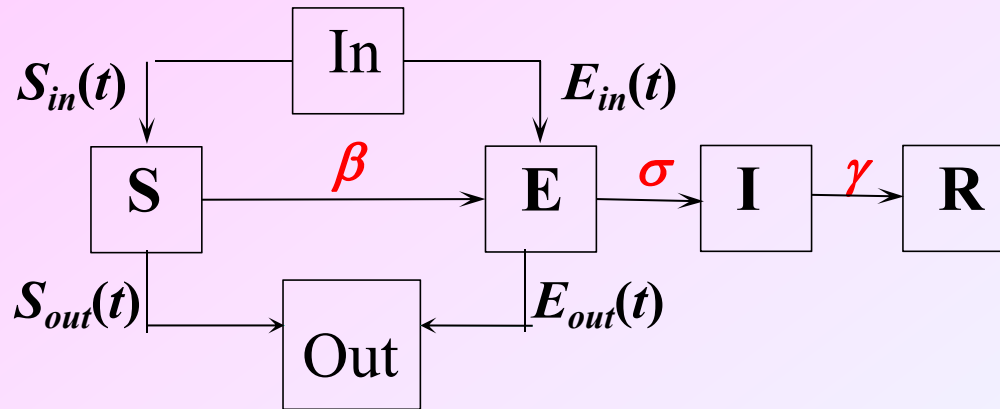
- The most recent epidemiological data based on daily COVID-19 outbreak numbers reported by the National Health Commission of China were retrieved

Situation report (in Chinese) 2020. Available online:
http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml

- Migration index based on the daily number of inbound and outbound events by rail, air and road traffic, were sourced from a web-based program

Baidu qianxi (in Chinese) 2020 Available online:
<https://qianxi.baidu.com/>

Estimation of parameters



$\beta = kb$, $k \sim$ people exposed to each day by infected.

$b \sim$ probability of transmission when exposed.

$\sigma \sim$ incubation rate, $1/\sigma \sim$ duration of incubation

$$\begin{aligned} \frac{dE}{dt} &= \frac{\beta SI}{N} - \sigma E \\ \frac{dI}{dt} &= \sigma E - \gamma I \end{aligned} \quad \Rightarrow ? \quad \frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I \approx (\beta - \gamma)I, \quad I_{t=0} = 1 \quad \Rightarrow \quad I = e^{(kb - \gamma)t}$$

• multiple fitting with data $\Rightarrow b = 0.05249, \gamma = 0.154$

• I will be quarantined. $k_1 = 3 \Rightarrow \beta_1 = k_1 b = 0.15747$

• E will have normal contact. $k_2 = 15 \Rightarrow \beta_2 = k_2 b = 0.78735$

• incubation period has been reported to be between 2 to 14 days, we chose the midpoint of 7 days. $\Rightarrow \sigma = 1/7$

Results

Epidemic progression in Hubei, Guangdong, Zhejiang and Mainland China

- Confirmed cases of in Hubei, Guangdong, Zhejiang on Feb 10 were 31728, 1177, 1117, representing **80% of total cases** nationwide.
- The **migration index out of Guangdong, Zhejiang** were greater than the inflow and were largest between Jan 7 and Jan 23.
- The **migration index into Hubei** was greater than the outflow before Jan 23, signaling the homeward return for Spring Festival.

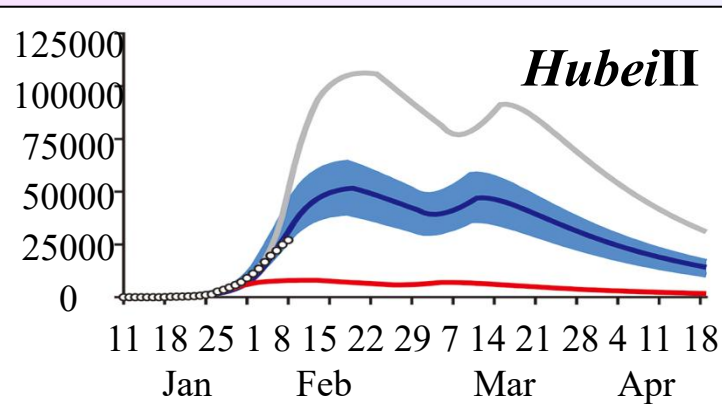
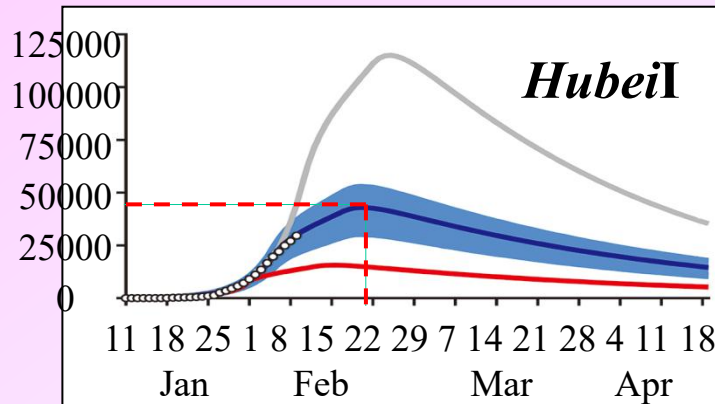
In Hubei, the model predicted a **epidemic peak on Feb 20 with 42792** and total case numbers reaching with 59578 nearing its end by late Apr.

If interventions were **delayed by five days**, a peak of 115061 cases would be reached by Feb 25 with total case numbers reaching 167598.

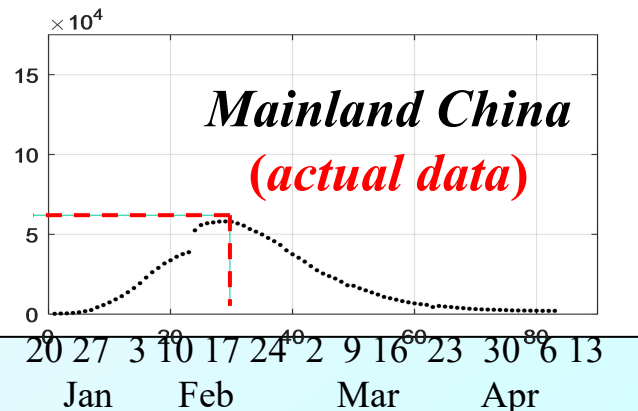
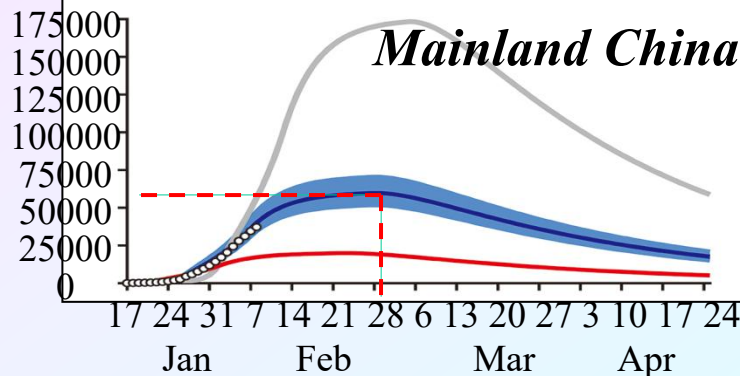
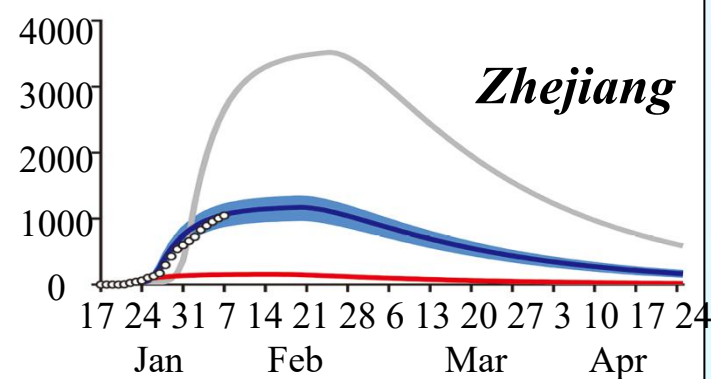
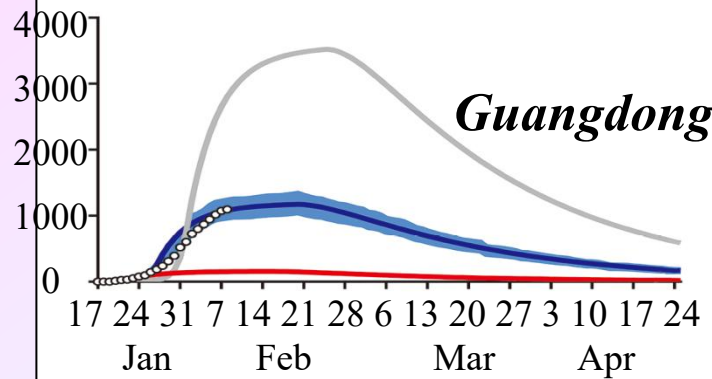
Results

when **interventions** were introduced on Jan 23 (blue),
five days later (grey) and five days earlier (red).

Number of active infections



$I \sim$ where
quarantine
ceased,
allowing
normal
migration.



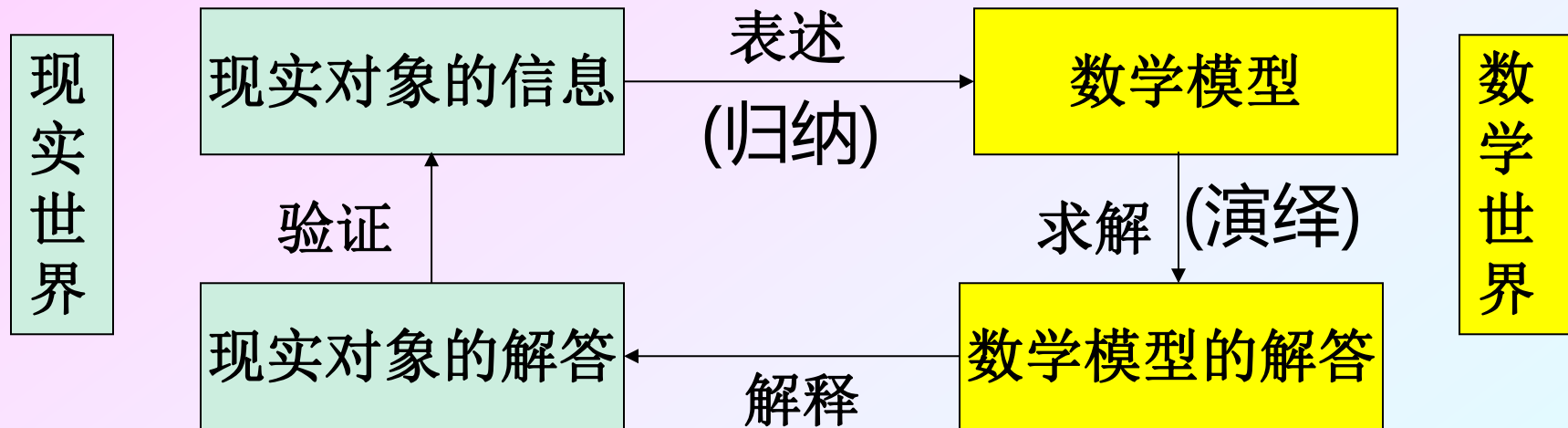
Discussion

- The actual epidemic trend since our analyses **has fit well with our predicted curve** (to 25 Feb).
- With the migrants beginning to return to Guangdong and Zhejiang, concerns spark over **potential increase in imported cases**.
- The continued policy of “early detection” and subsequent isolation **might be effective in preventing a second epidemic wave in Guangdong and Zhejiang**.
- We tested the model **sensitivity to different incubation time** and found that shorter incubation time will accelerate the epidemic peak and result in a smaller epidemic size.
- Our model did not account for other factors that may increase confirmed case numbers, such as **diagnostic capacity**.

Conclusions

- Our dynamic SEIR model was **effective** in predicting the COVID-19 epidemic peaks and sizes.
- The implementation of **control measures** on January 23 was predicted to **reduce** the COVID-19 epidemic size in China, and the policy of strict monitoring and early detection should **remain in place until the end of April 2020**.

数学建模——数学与现实之间的桥梁



表述 根据建模目的和信息将实际问题“翻译”成数学问题.

求解 选择适当的数学方法求得数学模型的解答.

解释 将数学语言表述的解答“翻译”回实际对象.

验证 用现实对象的信息检验得到的解答.

实践 \Rightarrow **理论** \Rightarrow **实践**

数学建模在科技发展和**社会进步**中的重要地位

- 作为用数学方法解决实际问题桥梁，
数学建模的历史与数学同样悠久。

欧几里德

《几何原本》

光反射定律

阿基米德

浮力定律

杠杆原理

伽利略

落体定律

惯性原理

牛顿

万有引力定律

微积分

数学建模不是什么新东西

- 为什么直到**20世纪中期**数学建模才逐渐得到普遍重视和广泛应用，并随之进入高等院校课堂呢？

数学建模在科技发展和**社会进步**中的重要地位

计算机技术迅速发展为数学建模提供了强有力的**工具**。

- 在一般工程技术领域数学建模仍然**大有用武之地**。
- 在高新技术领域，数学建模几乎是必不可少的手段——**数学科学**是**关键的、普遍的、可应用的**技术。
- 计算机技术与数学建模对于**数字经济**时代的科技发展可以比作“**如虎添翼**”。



数学建模在科技发展和**社会进步**中的重要地位

随着人类社会的不断进步以及众多**学科定量化、精准化**的需要，数学迅速进入经济、人口、生物、医学等领域，为数学建模开拓了许多新的**处女地**。

- 出现了如计量经济学、人口控制论、生物数学、地质数学等新的**交叉学科**。
- 用数学方法研究这些**非物理领域中的定量关系**，数学建模成为其首要的、关键的步骤。

教育必须适应**科技发展**的需要，顺应**社会进步**的潮流

数学建模引入大学教育促进了学生



综合素质的培养与提高

- 数学教育本质上是一种**素质教育**，要让学生领会到数学的精神实质和思想方法，成为得心应手的工具。
- 数学教育应培养两种能力：**算数学** (计算、推导、证明...) 和**用数学** (用 数学方法分析、解决实际问题)。
- **传统**的数学教学体系和内容偏重前者，忽略后者。
- 引入数学建模让学生参加数学用于实际的尝试,参与发现和创造的过程，更好地**应用、理解和热爱数学**。

数学建模进入大学课堂



- 20世纪60~70年代数学建模进入西方国家的大学。
- 1982年起复旦大学、清华大学等开设数学建模课。
- 目前大多数高校开设了各种形式、内容的建模课程。
- 姜启源编《数学模型》(高等教育出版社, 1987.4)是国内第1本建模教材, 现已出第五版, 总印数超过百万。



- 三十多年国内出版数学建模内容的教材超过200本。



《数学模型(第五版)》的 在线学习平台

ISBN: 978-7-89510-806-6
出版时间: 2020.6
出版单位: 高等教育出版社
高等教育电子音像出版社

- 本案例精讲以视频讲解的方式对《数学模型(第五版)》中的全部80个教学案例，进行深入细致地讲解与剖析，以帮助学生进一步提高数学建模的能力。

《数学模型(第五版)》的在线学习平台

全书购买 (60元)

单个购买 (3元/个)

第1章 建立数学模型

1.3 包饺子中的数学 (免费试看)

1.4 路障间距的设计 (免费试看)

1.5 椅子在不平的地面上放稳吗 (免费试看)

第2章 初等模型

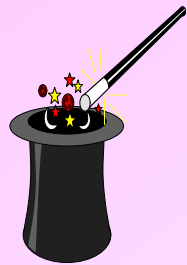
2.1 双层玻璃窗的功效 (免费试看)



扫码观看

双层玻璃窗的功效

清华大学 姜启源



怎样学习数学建模

数学建模与其说是一门**技术**，不如说是一门**艺术**

技术大致有章可循

艺术无法归纳成普遍适用的准则

- 艺术家的培养需要大量的**观摩**和前辈的**指教**，更离不开长期、艰苦的**亲身实践**。

培养数学建模的**意识和能力**

- 敏锐地**发现**日常生活中听到、看到、接触到的**可以、能够甚至需要**利用数学知识给以分析、解决的**实际问题**，并且**从数学建模的角度思考、研究**，力求得到一个基本满意的结果。

学习数学建模的**两条途径**

1. 学好他人的

学习、分析、评价乃至推广、改进他人的模型。

- 首先要**学懂**，然后**分析**为什么这样作，有什么优缺点，再**尝试**能不能作些推广、改进。

2. 做好自己的

亲自动手，踏实地完成几个实际问题的**建模过程**。

- 从小题目做起；作书上的题目不要先看解答。
- 积极参加各种层次的**数学建模竞赛**。

对两种常用建模方法的一些看法

注意层次分析方法的局限

- 层次分析主要基于相对度量，适用于尚无太多知识的新问题和模糊、抽象的准则，不宜用于已有充分了解的老问题和明确、具体的准则。

把握灰色模型预测的本质

- 灰色模型GM (1,1)本质上是实际数据作为指数序列来建模和预测，用最小二乘法估计参数。
- 对于不能确定来自指数序列的数据，不建议随便采用GM(1,1)来建模和预测。

注意层次分析方法的局限

- 成对比较法只是将定性比较加以量化、为了得到权重而采用的手段，如果权重本来可以定量地获得 (特别是方案对准则的权重)，就不要采用成对比较。
- 不要滥用成对比较法，把它作为个人主观赋权的工具 (把已经认定的权重用成对比较阵表示出来)。
- 准则对目标的权重宜采用客观性强的信息熵法，以便对那些辨别目标性能更强的准则 (按照方案对准则权重的区分性衡量) 赋予较大的权重。

注意层次分析方法的局限

- 对于由目标、准则、方案构成的层次结构的决策问题，**确定**准则对目标、方案对准则**权重**的方法可分为**相对度量**和**绝对度量**。**成对比较**属于前者，用定量尺度来描述方案或准则的特征属于后者。
- 一般来说，相对量测偏于主观、定性；绝对量测偏于客观、定量，应**尽量采用绝对度量**。
- **绝对量测**的另一优点：新方案加入或旧方案退出时原有方案的**结果不会改变**；用**相对量测**要重新做比较，原有方案的**结果可能改变**。

注意层次分析方法的局限

- 对已有充分了解的老问题和明确、具体的准则，建议采用绝对度量. 多属性决策 (Multiple Attribute Decision Making) 主要利用绝对度量.
- 层次分析法出现于20世纪70年代，80年代在我国迅速推广. 而比它早些年问世、可以处理同类实际问题的多属性决策 在国内的影响却不如前者.
- 层次分析法与多属性决策的建模过程及应用实例可参考姜启源 谢金星 叶俊编 《数学模型(第五版)》7.1,7.2节 (P233~246).

把握灰色模型预测的本质

GM(1,1)模型 $x^{(0)}(k) + ax^{(1)}(k) = b$

$x^{(0)}(k)$ 与 $x^{(1)}(k)$
仅相差常因子 a

$x^{(0)}(k) \geq 0, k = 1, \dots, n$ 原始序列 $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i)$ 累加序列

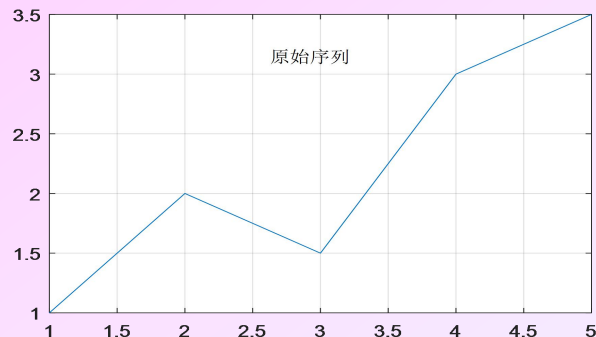
参数的最小二乘估计

$$\hat{a} = [a, b]^T = (B^T B)^{-1} B^T Y \quad Y = \begin{bmatrix} x^{(0)}(2) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}, B = \begin{pmatrix} -x^{(1)}(2) & 1 \\ \vdots & \vdots \\ -x^{(1)}(n) & 1 \end{pmatrix}$$

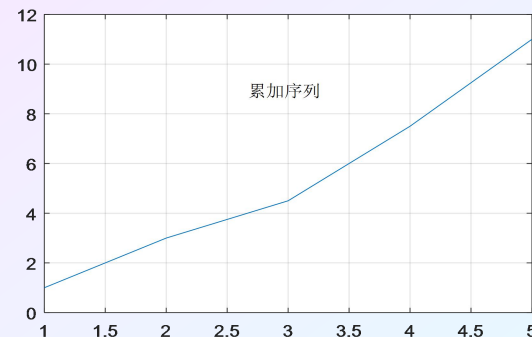
建模和预测 $\hat{x}^{(0)}(k) = (-a)(x^{(0)}(1) - \frac{b}{a}) (\frac{1}{1+a})^k$ 指数模型

- 序列累加相当于函数积分,函数与原函数相差常数因子的只有指数函数,故GM (1,1) 适用于指数序列.
- 采用GM (1,1)的前提: 原始数据为指数函数序列.

把握灰色模型预测的本质



序列累加减小
震荡利于建模



- 对指数函数GM (1,1)用原始与累加序列的线性性建模, 会比将原始序列作对数变换后进行线性拟合好吗?
- 对于其他函数序列, GM (1,1)的效果如何?
- 若对指数原始序列直接作非线性最小二乘拟合呢?
- 尚未见到理论上对这两种方法进行比较的论述, 可利用数值模拟对其建模和预测效果进行比较.

把握灰色模型预测的本质

数值模拟的设计

- 选择简单的指数函数、线性函数和二次函数，如
$$x_1(t) = 10e^{0.5t}, x_2(t) = t + 10, x_3(t) = 0.5t^2 + t + 10$$
- 取离散值并添加正态分布随机数 $\varepsilon(k)$ 产生原始序列
$$x_1^{(0)}(k) = 10e^{0.5k} + \varepsilon(k), x_2^{(0)}(k) = k + 10 + \varepsilon(k)$$
$$x_3^{(0)}(k) = 0.5k^2 + k + 10 + \varepsilon(k), \varepsilon(k) \sim N(0, \sigma^2)$$
- 对各序列取前 n 数据分别用GM(1,1) 和普通的线性拟合建模，并对第 $n+1$ 数据预测，计算误差。
- 对每个序列进行 m 次模拟，计算平均误差，比较两种方法得到的结果。

把握灰色模型预测的本质

数值模拟 ($n=5,10, m=10000$) 的初步结论

- 对指数函数添加随机误差产生序列，GM (1,1) 建模和预测的效果好于普通的线性拟合。
- 当随机误差变大时GM (1,1) 建模和预测的优势比线性拟合更加显著。
- 对线性函数和二次函数添加随机误差产生序列，GM (1,1)建模和预测效果不如普通的线性拟合。。
- 对指数函数GM(1,1)拟合效果优于线性最小二乘，与非线性最小二乘相比呢？——再作数值模拟！

- 作为指数序列参数拟合的方法之一， $GM(1,1)$ 的效果明显优于线性最小二乘，但是不如平常较少采用的非线性最小二乘。
- 如果按照常识或数据作图，拟合对象并非指数函数，不建议随便采用 $GM(1,1)$ 模型。
- 对于所谓“小数据、贫信息”的建模，具有先天不足的缺陷，不论选取什么拟合方法 (总是后天性的)，都难有实质性的改善。
- 积极开发大数据的建模，具有现实的重要性。

谢谢大家!