数学实验大作业

灰色系统中的预测实例

2023 年 5 月 28 日, *Nicolas-Keng*

目录

问题的提出与模型介绍

本问题节选自刘思峰《灰色系统理论及其应用》一书与蔡锁章《数学建模原理与方法》一书. 该问题在当下疫情防控措施解除, 旅游业与服务业反弹的时间仍有很大的参考价值, 可以帮助我们更好地备战国赛.

问题的提出

2003 年的 SARS 疫情对中国部分行业的经济发展产生了一定影响, 特别是对部分疫情较严重的省市的相关行业所造成的影响是显著的. 经济影响主要分为直接经济影响和间接影响. 直接经济影响涉及商品零售业, 旅游业, 综合服务等行业, 很多方面难以进行定量地评估; 现仅就 SARS 疫情较重的某市商品零售业, 旅游业和综合服务业的影响进行定量的评估分析.

已知某市从 1997 年 1 月到 2003 年 12 月的商品零售额, 接待旅游人数和综合服务收入的统计数据如下表:

表 1 商品的零售额（单位：亿元）

年代	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
1997	83.0	79.8	78.1	85.1	86.6	88.2	90.3	86.7	93.3	92.5	90.9	96.9
1998	101.7	85.1	87.8	91.6	93.4	94.5	97.4	99.5	104.2	102.3	101.0	123.5
1999	92.2	114.0	93.3	101.0	103.5	105.2	109.5	109.2	109.6	111.2	121.7	131.3
2000	105.0	125.7	106.6	116.0	117.6	118.0	121.7	118.7	120.2	127.8	121.8	121.9
2001	139.3	129.5	122.5	124.5	135.7	130.8	138.7	133.7	136.8	138.9	129.6	133.7
2002	137.5	135.3	133.0	133.4	142.8	141.6	142.9	147.3	159.6	162.1	153.5	155.9
2003	163.2	159.7	158.4	145.2	124.0	144.1	157.0	162.6	171.8	180.7	173.5	176.5

表 2 接待海外旅游人数（单位：万人）

年代	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
1997	9.4	11.3	16.8	19.8	20.3	18.8	20.9	24.9	24.7	24.3	19.4	18.6
1998	9.6	11.7	15.8	19.9	19.5	17.8	17.8	23.3	21.4	24.5	20.1	15.9
1999	10.1	12.9	17.7	21.0	21.0	20.4	21.9	25.8	29.3	29.8	23.6	16.5
2000	11.4	26.0	19.6	25.9	27.6	24.3	23.0	27.8	27.3	28.5	32.8	18.5
2001	11.5	26.4	20.4	26.1	28.9	28.0	25.2	30.8	28.7	28.1	22.2	20.7
2002	13.7	29.7	23.1	28.9	29.0	27.4	26.0	32.2	31.4	32.6	29.2	22.9
2003	15.4	17.1	23.5	11.6	1.78	2.61	8.8	16.2	20.1	24.9	26.5	21.8

表 3 综合服务业累计数额（单位：亿元）

年代	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
1997	96	144	194	276	383	466	554	652	747	832	972
1998	111	169	235	400	459	565	695	805	881	1011	1139
1999	151	238	335	425	541	641	739	866	975	1087	1238
2000	164	263	376	531	600	711	913	1038	1173	1296	1497
2001	182	318	445	576	708	856	1000	1145	1292	1435	1667
2002	216	361	504	642	818	979	1142	1305	1479	1644	1920
2003	241	404	584	741	923	1114	1298	1492	1684	1885	2218

图 1: 问题数据表格

试根据这些历史数据建立预测评估模型, 评估 2003 年 SARS 疫情给该市的商品零售业, 旅游业和综合服务业所造成的影响.

模型介绍

灰色系统理论是基于关联空间, 光滑离散函数等概念定义灰导数与灰微分方程, 进而用离散数据列建立微分方程形式的动态模型, 由于这是本征灰色系统的基本模型, 而且模型是近似的, 非唯一的, 故这种模型为灰色模型记为 GM(Grey Model), 即灰色模型是利用离散随机数经过生成变为随机性被显著削弱而且较有规律的生成数, 建立起的微分方程形式的模型, 这样便于对其变化过程进行研究和描述.

灰色预测是通过原始数据的处理和灰色模型的建立, 发现和掌握系统发展规律, 对系统的未来状态作出科学的定量预测. 目前应用较多的灰色预测模型是 GM$(1,1)$ 模型, 灰色马尔可夫预测模型等. 是指利用 GM 模型对系统行为特征的发展变化规律进行估计预测, 同时也可以对行为特征的异常情况发生的时刻进行估计计算, 以及对在特定时区内发生事件的未来时间分布情况做出研究等等. 这些工作实质上是将随机过程当作灰色过程, 随机变量当作灰变量, 并主要以灰色系统理论中的 GM$(1,1)$ 模型来进行处理.

灰色预测在工业, 农业, 商业等经济领域, 以及环境, 社会和军事等领域中都有广泛的应用, 特别是依据目前已有的数据对未来的发展趋势做出预测分析.

GM$(1,1)$ 模型

GM$(1,1)$ 的定义

设 $x^{(0)}$ 为 n 个元素的数列 $x^{(0)}=\left(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n)\right)$, $x^{(0)}$ 的 AGO 生成数列为

$$x^{(1)}=\left(x^{(1)}(1),x^{(1)}(2),\cdots,x^{(1)}(n)\right),\quad x^{(1)}(k)=\sum_{i=1}^{k}x^{(0)}(i)(k=1,2,\cdots,n),$$

则定义 $x^{(1)}$ 的灰导数为

$$\mathrm{d}(k)=x^{(0)}(k)=x^{(1)}(k)-x^{(1)}(k-1),$$

令 $z^{(1)}$ 为数列 $x^{(1)}$ 的紧邻均值数列, 即

$$z^{(1)}(k)=0.5x^{(1)}(k)+0.5x^{(1)}(k-1),\quad k=2,3,\cdots n,$$

则 $z^{(1)}=\left(z^{(1)}(2),z^{(1)}(3),\cdots,z^{(1)}(n)\right)$. 于是定义 GM$(1,1)$ 的灰微分方程模型为

$$\mathrm{d}(k)+az^{(1)}(k)=b,\quad \left(x^{(0)}(k)+az^{(1)}(k)=b\right),$$

其中 $x^{(0)}(k)$ 称为灰导数, a 称为发展系数, $z^{(1)}(k)$ 称为白化背景值, b 称为灰作用量. 将 $k=2,3,\cdots,n$ 代入上式中有

$$\left\{\begin{array}{l}x^{(0)}(2)+az^{(1)}(2)=b\\x^{(0)}(3)+az^{(1)}(3)=b\\\cdots\cdots\cdots\\x^{(0)}(n)+az^{(1)}(n)=b\end{array}\right.$$

令

$$Y=\left(x^{(0)}(2),x^{(0)}(3),\cdots,x^{(0)}(n)\right)^{T},\ u=(a,b)^{T},\ B=\begin{pmatrix}-z^{(1)}(2) & 1\\-z^{(1)}(3) & 1\\\vdots & \vdots\\-z^{(1)}(n) & 1\end{pmatrix}$$

称 Y 为数据向量, B 为数据矩阵, u 为参数向量, 则 GM$(1,1)$ 模型可以表示为矩阵方程 $Y=Bu$. 由最小二乘法可以求得

$$\hat{u}=(\hat{a},\hat{b})^T=\left(B^TB\right)^{-1}B^TY.$$

GM$(1,1)$ **的白化**

对于 GM$(1,1)$ 的上述灰微分方程, 如果将 $x^{(0)}(k)$ 的时刻 $k=2,3,\cdots,n$ 视为连续的变量 t, 则数列 $x^{(1)}$ 就可以视为时间 t 的函数, 记为 $x^{(1)}=x^{(1)}(t)$, 并让灰导数 $x^{(0)}(k)$ 对应于导数 $\frac{\mathrm{d}x^{(1)}}{\mathrm{d}t}$, 背景值 $z^{(1)}(k)$ 对应于 $x^{(1)}(t)$. 于是得到 GM$(1,1)$ 的灰微分方程对应的白微分方程为

$$\frac{\mathrm{d}x^{(1)}}{\mathrm{d}t}+ax^{(1)}=b,$$

称之为 GM$(1,1)$ 的白化型.

GM$(1,1)$ **预测值的检验**

建立模型 GM$(1,1)$, 则可以得到预测值

$$\hat{x}^{(1)}(k+1)=\left(x^{(0)}(1)-\frac{b}{a}\right)e^{-ak}+\frac{b}{a},\ k=1,2,\cdots,n-1,$$

且 $\hat{x}^{(0)}(k+1)=\hat{x}^{(1)}(k+1)-\hat{x}^{(1)}(k)$, $k=1,2,\cdots,n-1$.

令残差为 $\varepsilon(k)$, 计算

$$\varepsilon(k)=\frac{x^{(0)}(k)-\hat{x}^{(0)}(k)}{x^{(0)}(k)},\ k=1,2,\cdots,n,$$

如果 $\varepsilon(k)<0.2$, 则可认为达到一般要求; 如果 $\varepsilon(k)<0.1$, 则认为达到较高的要求.

另外, 也需首先由参考数据 $x^{(0)}(k-1),x^{(0)}(k)$ 计算出级比 $\lambda(k)$, 再用发展系数 a 求出相应的级比偏差

$$\rho(k)=1-\left(\frac{1-0.5a}{1+0.5a}\right)\lambda(k),$$

如果 $\rho(k)<0.2$, 则可认为达到一般要求; 如果 $\rho(k)<0.1$, 则认为达到较高的要求.

问题的解答

问题分析与模型假设

根据所掌握的历史统计数据可以看出, 在正常情况下, 全年的平均值较好地反映了相关指标的变化规律, 这样可以把预测评估分成两部分:

(1) 利用灰色理论建立 GM(1,1) 模型, 由 1997 – 2002 年的平均值预测 2003 年平均值;

(2) 通过历史数据计算每个月的指标值与全年总值的关系, 从而可预测出正常情况下 2003 年每个月的指标值, 再与实际值比较可以估算出 SARS 疫情实际造成的影响.

针对本文提出的问题, 给出下面两条模型假设:

(1) 假设该市的统计数据都是可靠准确的;

(2) 假设该市在 SARS 疫情流行期间和结束之后, 数据的变化只与 SARS 疫情的影响有关, 不考虑其它随机因素的影响.

建立灰色预测模型 GM(1,1)

由已知数据, 对于 1997-2002 年某项指标记为矩阵 $A=(a_{ij})_{6\times 12}$, 计算每年的年平均值, 记为

$$x^{(0)}=\left(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(6)\right),$$

并要求级比 $\lambda(i)=\dfrac{x^{(0)}(i-1)}{x^{(0)}(i)}\in(0.7515,1.3307)$, $(i=2,3,\cdots,6)$. 对 $x^{(0)}$ 作一次累加, 则:

$$x^{(1)}(1)=x^{(0)}(1),\quad x^{(1)}(i)=\sum_{k=1}^{i}x^{(0)}(k),\quad (i=2,3,\cdots,6).$$

记 $x^{(1)}=\left(x^{(1)}(1),x^{(1)}(2),\cdots,x^{(1)}(6)\right)$, 取 $x^{(1)}$ 的加权均值, 则

$$z^{(1)}(k)=\alpha x^{(1)}(k)+(1-\alpha)x^{(1)}(k-1),\quad (k=2,3,\cdots,6),$$

这里 α 为任一确定参数. 记 $z^{(1)}=\left(z^{(1)}(2),z^{(1)}(3),\cdots,z^{(1)}(6)\right)$.

于是取发展灰度 a 与内生控制灰度 b, 列出 GM(1,1) 的白化微分方程模型

$$\frac{\mathrm{d}x^{(1)}}{\mathrm{d}t}+ax^{(1)}=b.$$

由于 $x^{(1)}(k)-x^{(1)}(k-1)=x^{(0)}(k)$, 取 $x^{(0)}(k)$ 为灰导数, $z^{(1)}(k)$ 为背景值, 则上述模型相应的灰微分方程为

$$x^{(0)}(k)+az^{(1)}(k)=b,\quad (k=2,3,\cdots,6),$$

矩阵形式

$$Y=B(a,b)^T,\quad \begin{cases} Y=\begin{pmatrix} x^{(0)}(2) & x^{(0)}(3) & \cdots & x^{(0)}(6)\end{pmatrix} \\ B=\begin{pmatrix} -z^{(1)}(2) & -z^{(1)}(3) & \cdots & -z^{(1)}(6) \\ 1 & 1 & \cdots & 1\end{pmatrix}^T \end{cases}$$

用最小二乘法求得参数的估计值为

$$(\hat{a},\hat{b})^T=(B^TB)^{-1}B^TY,$$

于是上述灰微分方程有特解

$$\hat{x}^{(1)}(t+1)=\left(x^{(0)}(1)-\frac{b}{a}\right)e^{-at}+\frac{b}{a},$$

则

$$\hat{x}^{(0)}(k+1)=\hat{x}^{(1)}(k+1)-\hat{x}^{(1)}(k)=\left(x^{(0)}(1)-\frac{b}{a}\right)\left(e^{-ak}-e^{-a(k-1)}\right).$$

于是我们可以得到 2003 年的平均值为 $\bar{x}$, 则预测 2003 年的总值为 $Z=12\bar{x}$. 根据历史数据, 可以统计计算出 2003 年第 i 个月的指标值占全年总值的比例 u_i, 记 $u=(u_1,u_2,\cdots,u_{12})$, 于是可得 2003 年每一个月的指标值为 $V=Zu$.

模型求解

商品零售额

由数据表 1, 计算可得年平均值, 一次累加值分别为

$$\begin{cases} x^{(0)}=(87.6167,98.5,108.475,118.4167,132.8083,145.4083) \\ x^{(1)}=(97.6167,186.1167,294.5917,413.0083,545.8167,691.225) \end{cases}$$

显然 $x^{(0)}$ 的所有级比都在可容区域内, 经检验, 在这里取参数 $\alpha=0.4$ 比较合适, 于是

$$z^{(1)}=(127.0167,229.5067,341.9583,466.1317,603.98),$$

由最小二乘法求得 $a=-0.099, b=85.5985$, 2003 年的月平均值为 $\bar{x}=162.8793$ 亿元; 年总值为 $Z=1954.6$ 亿元; 每月的比例为

$$u=(0.0794,0.0807,0.0749,0.0786,0.0819,0.0818,0.0845,0.0838,0.0872,0.0886,0.0866,0.092),$$

故 2003 年 $1-12$ 月的预测值为

$$V=Zu=(155.2,157.7,146.4,153.5,160.1,159.8,165.1,163.8,170.5,173.1,169.3,179.8).$$

将预测值与实际统计值进行比较如下表所示:

月份	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
预测值	155.2	157.7	146.4	153.5	160.1	159.8	165.1	163.8	170.5	173.1	169.3	179.8
实际值	163.2	159.7	158.4	145.2	124.0	144.1	157.0	162.6	171.8	180.7	173.5	176.5

计算的 Matlab 程序如下:

```
clc,clear
load data1.txt  %把原始数据保存在纯文本文件data1.txt中
data1(end,:)=[];m=size(data1,2);
x0=mean(data1,2);
x1=cumsum(x0)
alpha=0.4;n=length(x0);
z1=alpha*x1(2:n)+(1-alpha)*x1(1:n-1)
Y=x0(2:n);B=[-z1,ones(n-1,1)];
ab=B\Y
k=6;
x7hat=(x0(1)-ab(2)/ab(1))*(exp(-ab(1)*k)-exp(-ab(1)*(k-1)))
z=m*x7hat
u=sum(data1)/sum(sum(data1))
v=z*u
```

接待海外旅游人数

由数据表 2, 类似地计算年平均值 $x^{(0)}$ 和一次累加值 $x^{(1)}$. 取参数 $\alpha = 0.5$, 并加权平均值 $z^{(1)}$. 最终求得 $a = -0.0938$, $b = 16.2671$, $\bar{x} = 30.2649$, $Z = 12\bar{x} = 363.1785$,

$$u = (0.0407, 0.0732, 0.0703, 0.0878, 0.0907, 0.0848, 0.0836, 0.1022, 0.101, 0.1041, 0.0914, 0.0701).$$

于是可得到 2003 年的接待海外旅游人数的预测值, 并与实际值比较如下表所示:

月份	1 月	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
预测值	14.8	26.2	25.5	31.9	33.0	30.8	30.4	37.1	36.7	37.8	33.2	25.5
实际值	15.4	17.1	23.5	11.6	1.78	2.61	8.8	16.2	20.1	24.9	26.5	21.8

Matlab 程序与前程序完全一致, 这里不赘述.

综合服务业累计数据

由数据表 2, 类似地计算年平均值 $x^{(0)}$ 和一次累加值 $x^{(1)}$. 取参数 $\alpha = 0.5$, 并加权平均值 $z^{(1)}$. 最终求得 $a = -0.1361$, $b = 487.7639$, $\bar{x} = 1171.3$, $Z = 11\bar{x} = 12885$,

$$u = (0.0191, 0.031, 0.0433, 0.0591, 0.0728, 0.0875, 0.1046, 0.1205, 0.1358, 0.1515, 0.1749).$$

于是可得到 2003 年的综合服务业累计数额的预测值, 并与实际值比较如下表所示:

月份	2 月	3 月	4 月	5 月	6 月	7 月	8 月	9 月	10 月	11 月	12 月
预测值	249	399	558	762	938	1127	1348	1553	1750	1952	2254
实际值	241	404	584	741	923	1114	1298	1492	1684	1885	2218

Matlab 程序与前程序完全一致, 这里不赘述.

结果分析

根据该市的统计报告显示, 2003 年 4, 5, 6 三个月的实际商品零售额分别为 145.2, 124, 144.1 亿元. 在这之前, 根据统计部门的估计, 4, 5, 6 三个月份 SARS 疫情对该市的商品零售业的影响最为严重, 这三个月估计大约损失 62 亿元左右. 从我们的模型预测结果来计算, 4, 5, 6 三个月的损失为 60.1 亿元, 基本与专家的估计值相符, 8 月基本恢复正常, 这也说明了模型的正确性和可靠性.

旅游业是受影响最严重的行业之一, 最严重的 4 – 7 四个月就损失 100 多万人, 按最新统计数据, 平均每人消费 1002 美元计算, 大约损失 10 亿美元. 全年大约损失 162 万人, 约合 16.2 亿美元, 到年底基本恢复正常.

对于综合服务业中的部分行业影响较大, 如航空交通运输, 宾馆餐饮等, 但有些行业影响不大, 如电信通讯等, 总平均来看, 影响不算太大, 5 – 8 四个月大约损失 70 亿元.

该模型虽是就某经济指标的发展规律进行评估预测而建立的, 但类似地也适用于其它方面的一些数据规律的评估预测问题, 即该模型具有很广泛的应用性.