

数列灰预测在旅游统计与预测上的应用 ——以江苏省为例

罗朝霞 张树夫

(南京师范大学旅游系 210097 江苏 南京)

[摘要] 灰色理论自 20 世纪 70 年代提出以后 , 被广泛应用于社会、经济、科技、农业、生物等各个领域。灰色预测是灰色理论的主要内容之一 , 其中数列灰预测是灰色预测的主要类型之一。以江苏省国内游客量和旅游收入(1995 ~ 2001 年) 为原始数据建立数列灰预测模型 , 验证了这一预测模型的可信度 , 分析讨论了此种预测方法在旅游统计与预测上应用的优缺点及应用价值。

[关键词] 数列灰预测 游客量 旅游收入 江苏省

[中图分类号] F590 , [文献标识码] A , [文章编号] 1001-4616(2005)02-0117-05

The Application of Sequence Grey Prediction in Tourism Statistics and Prediction ——Jiangsu Province for Example

Luo Zhaoxia , Zhang Shufu

(Department of Tourism , Nanjing Normal University , 210097 , Nanjing , China)

Abstract Since grey theory was presented in 70s ' 20 century , it has been applied in many fields such as society , economy , science and technology , agriculture , and biology. Grey prediction is one of the main ideas of grey theory. The sequence grey prediction is one of major type of grey prediction. Based on the original data of national visitors and tourism receipts (from 1995 to 2001) in Jiangsu province , the prediction model is built , the accuracy of the prediction is verified , and in the end the advantages , disadvantages and application value of applying this method on tourism statistics and prediction are analyzed and discussed.

Key words sequence grey prediction , domestic visitors , tourism receipts , Jiangsu province

0 引言

灰色理论自 20 世纪 70 年代被提出后 , 现已广泛应用于各种领域^[1] , 旅游方面也有一些学者利用和研究此种理论^[2-7]。灰色理论主要应用于旅游产业结构的灰色关联度分析、旅游资源评价以及旅游收入和游客量预测 , 其中灰色预测主要用了数列灰预测技术。但以往的研究者们主要利用此种预测技术来得出一些数据为自己的研究项目所用 , 很少讨论为何可以应用此种预测方法 , 以及使用此种预测的前提、注意事项、优缺点等。

灰色预测包括数列灰预测、灾变灰预测、季节灾变灰预测等多种预测方法 , 其中数列灰预测是最普遍 , 应用最广泛的一种预测方法。GM(1 , 1) 模型是最简单的数列灰预测模型 , 它是一元线性回归模型的演化 , 是对时间序列变量的预测。具体地说 : 灰预测是现有状态向未来延伸的预测 , 为了获得可信的结果 , 要求建模数据具有全信息性。在旅游预测中 , 以每年国内游客量和旅游收入作为原始数据 , 每年的国内游客量和旅游收入是当年各种因子如 经济、政治、环境、生态、交通等共同作用的结果。显然影响当年国内游客量和旅游收入的因子很多 , 因此旅游因子具有灰信息覆盖 , 而当年的国内游客量和旅游收入是具体的、确定的 , 具有白信息覆盖 , 因此从旅

收稿日期 : 2004-12-17.

基金项目 : 江苏省教育厅自然科学基金资助项目(2003KJD170111)。

作者简介 : 罗朝霞 , 女 , 1979— , 硕士研究生 , 主要从事旅游规划与管理的学习与研究。E-mail : zhaozhao115@ sina. com

通讯联系人 : 张树夫 , 1949— , 教授 , 主要从事旅游消费行为与旅游规划的教学与研究。E-mail : zhangshufu49@ sina. com

万方数据

游因子到游客量或旅游收入符合灰因白果律. 根据《灰理论基础》定理可知: 游客量或旅游收入都是全部因子作用的结果, 它们的数据符合全信息性, 可以作为原始数据建立灰色预测模型.

1 建模分析

1.1 建模测算

以 1995 到 2001 年江苏省国内游客量和旅游收入分别作为原始数据建立数列灰预测模型. 其基本数据见表 1 所示.

(1) 设 1995 ~ 2001 年江苏省国内游客量原始数列 $x^{(0)}$ 为:

$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), x^{(0)}(4), x^{(0)}(5), x^{(0)}(6), x^{(0)}(7)) =$
 $(4.40, 4.86, 5.58, 6.12, 6.60, 7.19, 8.07),$

1995 ~ 2001 年江苏省国内旅游收入原始数列 $y^{(0)}$ 为:

$y^{(0)} = (y^{(0)}(1), y^{(0)}(2), y^{(0)}(3), y^{(0)}(4), y^{(0)}(5), y^{(0)}(6), y^{(0)}(7)) =$
 $(2.61, 3.00, 3.65, 4.42, 5.11, 5.88, 6.76).$

表1 江苏省1995 ~ 2001年国内游客量和旅游收入情况一览表

	1995 年	1996 年	1997 年	1998 年	1999 年	2000 年	2001 年
国内游客量 / (千万人次)	4.40	4.86	5.58	6.12	6.60	7.19	8.07
旅游收入 / (百亿元人民币)	2.61	3.00	3.65	4.42	5.11	5.88	6.76

资料来源: 江苏省统计信息网 <http://www.jssb.gov.cn>

(2) 对原始数列进行级比检验:

$\sigma_1(k) = x^{(0)}(k-1)/x^{(0)}(k) = (0.905, 0.871, 0.912, 0.927, 0.918, 0.891),$
 $\sigma_2(k) = y^{(0)}(k-1)/y^{(0)}(k) = (0.870, 0.822, 0.826, 0.867, 0.869, 0.869),$
 $\sigma_1(k) \in (1-0.129, 1), \sigma_2(k) \in (1-0.178, 1).$

已知 $n = 7$, 有界区 $(e^{-2/n+1}, e^{2/n+1}) = (0.7789, 1.1284) \subset (0.1353, 7.3890),$
级比区 $\sigma_1(k) \in (1-0.129, 1) \subset (0.7789, 1.1284),$
 $\sigma_2(k) \in (1-0.178, 1) \subset (0.7789, 1.1284).$

由以上可知: 级比区在级比可容区和级比界区内, 所以以上原始数据推断可以获得精度较高的 GM(1, 1) 模型.

(3) 对原始数列 $x^{(0)}$ 进行一次累加生成得到序列 $x^{(1)}$, 再对序列 $x^{(1)}$ 进行均值化生成得到数列 $z^{(1)}$.

$x^{(1)} = \text{AGO } x^{(0)},$
 $x^{(1)}(k) = x^{(1)}(k-1) + x^{(0)}(k) = (4.40, 9.26, 14.84, 20.96, 27.56, 34.75, 42.82),$
 $k = (1, 2, 3, 4, 5, 6, 7),$
 $z^{(1)} = \text{MEAN } x^{(1)},$
 $z^{(1)}(k) = 0.5 x^{(1)}(k-1) + 0.5 x^{(1)}(k) = (6.83, 12.05, 17.90, 24.26, 31.16, 38.79),$
 $k = (2, 3, 4, 5, 6, 7).$

(4) 建立 GM(1, 1) 模型.

$dx^{(1)}/dt + a x^{(1)} = b,$ 求解微分方程得到:

$$\hat{x}^{(1)}(k+1) = (x^{(0)}(k) - b/a)e^{-ak} + b/a \tag{1}$$

再还原便得到

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \tag{2}$$

(5) 用最小二乘法求解系数 $a, b.$

$$C = \sum_{k=2}^7 z^{(1)}(k) = 131.07, \quad D = \sum_{k=2}^7 x^{(0)}(k) = 38.42,$$
$$E = \sum_{k=2}^7 x^{(0)}(k) z^{(1)}(k) = 907.17, \quad F = \sum_{k=2}^7 z^{(1)}(k)^2 = 3576.42,$$
$$a = \frac{CD - (n-1)E}{(n-1)F - C^2} = \frac{CD - 6E}{6F - C^2} = -0.095, \quad b = \frac{DF - CE}{(n-1)F - C^2} = \frac{DF - CE}{6F - C^2} = 4.32, \quad b/a = -45.47.$$

(6) 将以上数据代入(1)、(2)两式得出预测值 $\hat{x}^{(1)}(k+1)$ 和 $\hat{x}^{(0)}(k+1)$, 即 GM(1,1) 的白化响应式为:
 $\hat{x}^{(1)}(k+1) = 49.87 e^{0.095k} - 45.47.$

(7) 残差检验.

相对残差为 $e^{(0)}(k) = \frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)} 100\%$, 平均残差 $e^{(0)} = \sum_{k=1}^7 |e(k)| / 7.$

精度 $p^{(0)} = (1 - e^{(0)})\%.$

(8) 对比列表.

从以上结论得出江苏省国内游客量预测值 $\hat{x}^{(0)}$, 并与原始值 $x^{(0)}$ 进行对比, 按上述同样步骤算出江苏省旅游收入预测值 $\hat{y}^{(0)}$, 通过对比得出此种预测方法的相对残差 $e^{(0)}(k)$, 具体数据见表2. 通过表2 算出精度 $p^{(0)}$ 分别为 98.42%、99.66%, 由此可见用此模型测出来的精确度较高.

表2 1995 ~ 2001 年江苏省国内游客量灰预测值和旅游收入灰预测值对比表

	$x^{(0)}$	$\hat{x}^{(0)}$	$e^{(0)}(k)$	$y^{(0)}$	$\hat{y}^{(0)}$	$e^{(0)}(k)$
1995 年	4.4	4.4	0%	2.61	2.61	0%
1996 年	4.86	4.987	-2.61%	3.00	3.14	-4.7%
1997 年	5.58	5.436	2.58%	3.65	3.67	-0.55%
1998 年	6.12	6.037	1.36%	4.42	4.28	3.16%
1999 年	6.6	6.483	1.77%	5.11	5.01	1.96%
2000 年	7.19	7.38	-2.64%	5.88	5.84	0.68%
2001 年	8.07	7.98	1.12%	6.76	6.82	-0.89%

1.2 建模分析

GM(1,1) 的定义型为

$$x^{(0)}(k) + a z^{(1)}(k) = b x^{(0)}(k) - 0.095 z^{(1)}(k) = 4.32 z^{(1)}(k) = 0.5 x^{(1)}(k-1) + 0.5 x^{(1)}(k).$$

根据上述模型的测算方法可以推断出 2002 年、2003 年江苏省国内游客量和旅游收入 $x^{(0)}(8)$ 、 $x^{(0)}(9)$ 、 $y^{(0)}(8)$ 、 $y^{(0)}(9)$.

$$x^{(0)}(8) - 0.095 z^{(1)}(8) = 4.32 \tag{3}$$

$$z^{(1)}(8) = 0.5 x^{(1)}(8-1) + 0.5 x^{(1)}(8) = 21.41 + 0.5 x^{(1)}(8) \tag{4}$$

由(3)、(4)式可以得出 $x^{(0)}(8) = 8.806$, 同理可以推出 $x^{(0)}(9) = 9.68$, $y^{(0)}(8) = 7.42$, $y^{(0)}(9) = 8.67$.

对比江苏省 2002 年、2003 年实际国内游客量和旅游收入列表3. 由表3 可知: 用灰色预测模型预测临近时间内的数据较为准确, 但随着时间的发展误差率越来越大, 其使用价值降低, 所以用此种方法预测未来某近段时间的游客量和旅游收入是比较准确的.

结合原始数列与预测值进行对比分析(见表3, 图1), 可以看出预测曲线和实际曲线比较接近, 说明该模型的可信度比较高, 曲线后面出现一点偏差是因为 2003 年的突发事件: 非典, 影响了旅游业的正常运转. 但是同样可以看出用模型预测的精确度比对模型本身检验的精确度要低, 通过该预测曲线仅可以大致了解江苏省国内游客量与旅游收入的大致发展趋势, 为旅游行业的发展提供一点参考信息. 当然可以通过数学方法进一步探讨曲线的平滑趋势与发展方向, 得到更精确的预测值和发展轨迹.

表3 2002、2003 年江苏省国内游客量和旅游收入实际值与预测值的对比

	实际国内游客量 / (千万人次)	预测国内游客量 / (千万人次)	精确度 / %	实际旅游收入 / (百亿元人民币)	预测旅游收入 / (百亿元人民币)	精确度 / %
2002 年	10.038	8.806	87.7	8.30	7.42	89.4
2003 年	11.4	9.68	84.9	9.75	8.67	88.9

资料来源: 江苏省统计信息网 <http://www.jssb.gov.cn/>

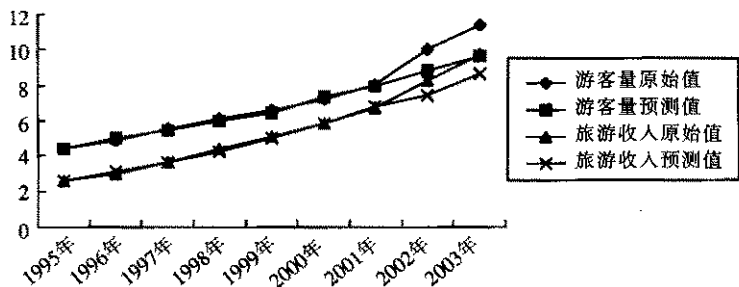


图1 1995~2003年江苏省国内游客量和旅游收入对比图

王海鸿在《中国入境旅游的灰色预测》中通过灰色数列模型预测2001~2005年中国入境游客量和旅游外汇收入年增长率^[2],结合2001~2003年中国实际入境游客量和旅游外汇收入增长率进行对比得出表4。由表4可知2001年、2003年的预测值误差较大,精确度有的甚至出现负数,这是由于2001年的9·11事件和2003年的非典两件突发事件的影响,使得中国在2001年和2003年的入境游客量和旅游外汇收入受到严重创伤,旅游业的正常运作受到破坏,致使实际值与预测值相差较大,有的误差率甚至高达300%多。由此可见,数列灰预测模型对于突发性环境变化的数据预测几乎没有什么优势。

表4 2001~2005年入境旅游发展预测与实际入境旅游发展状况对比表

入境游客量预测增长率		入境游客量实际增长率	精确度	旅游外汇收入预测增长率		旅游外汇收入实际增长率	精确度
2001 年	15. 1	6. 7	- 25. 4	14. 2	9. 7	53. 61	
2002 年	15. 7	9. 7	38. 2	14. 5	14. 57	99. 99	
2003 年	16. 2	- 6. 38	- 253. 9	15. 0	- 14. 61	- 102. 7	
2004 年	16. 0			13. 3			
2005 年	15. 4			12. 7			

1.3 模型注意事项

原始序列 $x^{(0)}(k)$ 通过一次累加生成转化为数列 $x^{(1)}(k)$,建立灰色预测模型其形式为:

$$dx^{(1)}/dx + ax^{(1)} = b \tag{5}$$

将(5)式转化为

$$x^{(0)}(k) + az^{(1)}(k) = b \tag{6}$$

数列 $x^{(1)}(k)$ 是数列 $x^{(0)}(k)$ 的累加生成,数列 $z^{(1)}(k)$ 是数列 $x^{(1)}(k)$ 的均值化生成,通过矩阵计算、最小二乘法等求出一级参数包、二级参数包、发展系数 a 和灰色作用量 b ,再用 a 、 b 代入(6)式求出预测算出的 $\hat{x}^{(0)}(k)$,比较测算值 $\hat{x}^{(0)}(k)$ 与原始值 $x^{(0)}(k)$,中间的误差率很小,说明灰微分方程(6)式非常接近真正的微分方程(5)的。所以可以用灰微分方程(6)式来预测未来发展状况。

通过以上分析可知应用此种模型应注意以下4个前提:

(1) 建模原始数据要求至少4个,并且具有 $X^{(1)}$ ($X^{(1)}$ 为 $x^{(1)}$ 上的等间隔子序列集)上信息浓度最大的灰导数。

(2) 并非所有 GM(1,1) 模型都是有效的,对于测算出值 $\hat{x}^{(0)}(k)$ 小于零、等于零或趋向无穷,以及 $d(k)$ 超过允许值范围的都是无效的,发展系数 a 必须属于 $(-2, 2)$ 区间才有效。

(3) 在开始计算前必须进行级比检验。根据灰色理论可知,级比可容区为 $\sigma(k) = e^{-a}$ 即可容区为 $(e^{-2}, e^2) = (0.1353, 7.3890)$,级比界区为 $(e^{-2/n+1}, e^{2/n+1})$,只有原始数列满足级比可容区和级比界区时模型才有效。

(4) 凡是从 GM(1,1) 白化模型及其白化响应式得到的结果只有与灰模型 GM(1,1) 不矛盾时才有价值。2001年、2003年的中国入境游客量和旅游外汇收入的原始数据不满足“灰导数白化背景值位于单调增的背景集中,与灰导数成份满足平射的数据非唯一”^[1]这一灰色建模的品质条件,所以白化模型测出的值没有什么价值。

2 讨论

2.1 模型的优势

(1) 回归模型、差分模型、时间序列模型都属于大样本量模型,而数列灰预测模型属于少数据模型,建立一

个常用的灰模型 $GM(1,1)$ 可允许数据少至 4 个^[1], 采数据较为节便. 对于某些旅游景点、旅游省市地区统计数据较少的单位可以采用此方法进行预测.

(2) 灰模型是既具差分又具微分性质的模型. 模型在关系上、性质上、内涵上具有不确定性. 即具有一定的灵活性, 更加符合现实情况^[1].

(3) 灰模型对于预测背景相似的数据有相对较高的精确度.

(4) 时间序列预测是用时间序列的几何特征和统计规律, 对序列未来发展进行估计, 是一种历史的和静态的描述和研究, 而数列灰预测是一种动态的和现实的分析与展望^[8], 旅游是一个动态系统, 所以这一预测模型更适用于旅游预测.

2.2 模型的劣势

(1) 由于游客量和旅游收入均为多因子共同作用的结果, 其背景不可能每年都一样, 例如 2001 年的 9·11 事件以及 2003 年的非典致使中国国际旅游业严重受挫, 许多地方游客量和旅游收入直线下降, 旅游业出现负增长, 因此实际数据与预测数据相差很大, 模型就失去其应有的价值. 故模型对于突发事件背景下的预测意义不大.

(2) 数列灰预测计算量大, 要求具有一定的系统设备. 此模型用计算机进行测算时要求计算机的配置较高, 况且一般的单位都很少安装灰色系统软件, 故目前普及率不高.

(3) 此类模型仍以历史数据为依据, 但由于现今旅游统计工作的不完善和统计技术的不完备等原因使得很多统计数据都不甚精确, 难免会影响预测结果的精确性.

(4) 数列灰预测模型建立后不是通用的, 只有对临近时间段的预测准确率才较高, 越往后发展其误差可能越大, 并且预测比较简单.

3 结论

综上所述, 数列灰预测可以在旅游统计与预测上应用, 但一旦环境及背景条件发生了变化, 或旅游地生命周期阶段发生了转变则应更换模型, 所以此种预测技术应从实际出发正确选用不能生搬硬套, 作为一种应用技术, 在环境背景相似数据又较少时是可行的. 另外, 数列灰预测模型不仅只有 $GM(1,1)$ 模型, 还有 $GM(2,1)$ 模型、VHST 数列预测模型等.

随着旅游业的发展, 旅游预测技术也不断丰富, 时间序列法、回归模型法、特尔菲法、引力模型等等, 各种方法各有所长、各有所短, 数列灰预测技术只是其中一种方法而已, 希冀旅游企业、旅游主管部门、旅游策划者可根据企业、市场与自身状况选择一种或多种比较适宜的预测方法, 以求统计和预测结果更为可信, 更具科学意义.

[参考文献]

- [1] 邓聚龙. 灰理论基础[M]. 武汉: 华中科技大学出版社, 2002. 1: 210—239.
- [2] 王海鸿. 中国入境旅游的灰色预测[J]. 天水行政学院学报, 2001 (3): 10—13.
- [3] 杨军强. 国际来华旅游人数动态预测[J]. 纺织高校基础科学学报, 2001, 14(4): 363—372.
- [4] 刘勇, 赵艺学. 山西省旅游业发展的灰色关联动态分析[J]. 山西大学学报(自然科学版), 2003, 26(1): 91—94.
- [5] 陈雪琼, 任晓春. 福建省旅游产业结构研究及其优化对策[J]. 华侨大学学报(哲社版), 2001 (3): 20—26.
- [6] 鲁小珍, 胡海波. 宁夏沙湖风景区的旅游市场分析[J]. 南京林业大学学报(人文社科版), 2002, 2(4): 61—63.
- [7] 汪华斌, 李江风. 鄂西清江三峡旅游资源多层次灰评价[J]. 长江流域资源与环境, 1999, 8(3): 261—269.
- [8] 王学萌. 灰色系统分析及实用计算程序[M]. 武汉: 华中科技大学出版社, 2001. 110.

[责任编辑: 丁蓉]