

马尔可夫链在太湖流域梅雨期预报中的应用

张宗国

(河海大学理学院 , 江苏 南京 210098)

摘要 :依据 1954 ~ 1991 年太湖流域的入、出梅日期以及梅雨集中期 ,应用马尔可夫链预报了太湖流域 1992 年和 1993 年的梅雨期 ,并利用其平稳分布分析了梅雨期系统在各状态下的概率 .应用结果表明 :在太湖流域气候条件稳定的情况下 ,历史数据越多越精确 ,预测也就越可靠 ;以马氏链预测梅雨状态的方法是一种概率预报方法 ,需通过大量的预测检验才能显示出该方法的准确性和可靠性 .
关键词 :太湖流域 ;梅雨期预报 ;马尔可夫链 ;随机矩阵 ;转移概率
中图分类号 :O21 **文献标识码** :A **文章编号** :1000-1980(2006)01-0116-03

太湖流域属于北亚热带海洋性湿润气候地带 ,梅雨明显且多受台风影响 ,历年 9 月份降雨量约 700 mm ,常年 6 月中旬入梅 .如果梅雨与 4、5 月间的春雨相继发生 ,则常会造成严重的洪涝灾害^[1] .由于洪涝灾害发生的程度与梅雨期的长短有明显的关系 ,所以研究梅雨期的长短具有很实际的意义^[1] .为此 ,本文依据 1954 ~ 1991 年太湖流域的入、出梅日期以及梅雨集中期 ,应用马尔可夫链预报了太湖流域 1992 年和 1993 年的梅雨期 ,并利用其平稳分布分析了梅雨期系统在各状态^[2]下的概率 .

1 理论基础

1.1 马尔可夫链

定义 1 设有随机过程 $\{X_n, n \in T\}$, $T = \{0, 1, 2, \dots\}$, 状态空间 $I = \{0, 1, 2, \dots\}$, 若对任意的 $m, n \in T$ 和任意的 $i_0, i_1, \dots, i_n, i_{n+m} \in I$, 有

$$p\{X_{n+m} = i_{n+m} \mid X_0 = i_0, X_1 = i_1, \dots, X_n = i_n\} = p\{X_{n+m} = i_{n+m} \mid X_n = i_n\}$$

则称 $\{X_n, n \in T\}$ 为马尔可夫链 , 简称马氏链 . 条件概率 $p\{X_{n+1} = j \mid X_n = i\}$ 称为马尔可夫链 $\{X_n, n \in T\}$ 在时刻 n 时的 1 步转移概率 , 简称为转移概率 , 记为 $p_{ij}(n; 1)$; $p\{X_{n+m} = j \mid X_n = i\}$ 称为时刻 n 时的 m 步转移概率 , 记为 $p_{ij}(n; m)$. 若 $p_{ij}(n; m) = p_{ij}(m)$, 特别地 , $p_{ij}(n; 1) = p_{ij}$, 则称此链是齐次的 , 其中 $i, j \in I$. 应用上主要研究齐次马尔可夫链^[3] .

1.2 转移概率矩阵

定义 2 1 步转移概率 p_{ij} 组成的矩阵 $P = \begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \ddots & \vdots \\ p_{m1} & \cdots & p_{mn} \end{bmatrix}$ 称为马氏链 $\{X_n, n \in T\}$ 的 1 步转移概率矩阵 .

它具有性质 (a) $p_{ij} \geq 0, i, j \in I$; (b) $\sum_{j \in I} p_{ij} = 1, i \in I$. 也可以定义 m 步转移概率矩阵 , 满足性质 (a) (b) 的矩阵称为随机矩阵^[4] .

1.3 平稳分布

定义 3 若

$$\pi_j = \sum_{i \in I} \pi_i p_{ij} \quad \sum_{j \in I} \pi_j = 1 \quad (\pi_j \geq 0)$$

则概率分布 $\{\pi_j, j \in I\}$ 称为马尔可夫链的平稳分布 .

2 数学模型的建立

根据一年内梅雨期的长短,可将各年的梅雨分为空梅、轻梅、常梅和重梅.其标准是:空梅,梅雨期小于或等于3d;轻梅,梅雨期小于或等于10d;重梅,梅雨期大于或等于25d;常梅,梅雨期在10~25d之间.在梅雨状态变化以a为单位的时间序列中,可以认为梅雨将来的等级在现在的等级已知的条件下与过去的等级无关,即满足马尔可夫性.于是,可以得到具

表 1 太湖流域梅雨历史资料^[1]

以上4个状态的马尔可夫链 $\{X_n, n \in T\}$.若用1,2,3,4来表示梅雨的4种等级,则状态空间 $I = \{1, 2, 3, 4\}$.当样本容量足够大时,1步转移概率可以近似地用状态相互转移的频率来描述^[1],即 $p_{ij} = M_{ij}/M_i, i, j \in I$.其中: M_i 为处于状态*i*的原始状态样本数; M_{ij} 为样本中由状态*i* 1步转移到状态*j*的样本数.于是可以由样本序列 x_n 获得近似的1步转移概率矩阵*P*.根据切普曼-柯尔莫哥洛夫方程,还可以确定出多步转移概率矩阵^[5].

Table 1 Historical plum rain information of Taihu Lake Basin

年份	入梅日期 (月-日)	出梅日期 (月-日)	梅雨 期/d	等级	年份	入梅日期 (月-日)	出梅日期 (月-日)	梅雨 期/d	等级
1954	06-01	08-02	62	重	1973	06-16	06-29	13	常
1955	06-27	07-08	11	常	1974	06-10	07-18	38	重
1956	06-05	07-19	44	重	1975	06-17	07-16	29	重
1957	06-14	07-09	25	常	1976	06-16	07-16	30	重
1958			0	空	1977	06-17	07-01	14	常
1959	06-28	07-07	9	轻	1978			0	空
1960	06-13	06-25	12	常	1979	06-18	07-24	36	重
1961	06-07	06-14	7	轻	1980	06-09	07-21	42	重
1962	06-17	07-07	20	常	1981	06-22	07-03	11	常
1963	06-22	07-08	16	常	1982	07-07	07-27	20	常
1964	06-24	06-27	3	空	1983	06-19	07-24	35	重
1965	06-25	06-27	2	空	1984	06-12	07-06	24	常
1966	06-13	07-12	29	重	1985	06-21	07-07	16	常
1967	06-24	07-09	15	常	1986	06-12	07-06	24	常
1968	06-23	07-11	8	常	1987	06-21	07-09	18	常
1969	06-24	07-16	22	常	1988	06-11	06-29	18	常
1970	06-18	07-18	30	重	1989	06-10	07-13	33	重
1971	05-26	06-23	28	重	1990	06-18	07-04	16	常
1972	06-20	07-03	13	常	1991	05-19	07-13	55	重

3 梅雨期的预测

表1是1954~1991年历年的梅雨起止日期及梅雨期,从时间上可以看出梅雨期的等级分别为重梅、常梅、轻梅和空梅.

从表1可以得出每年太湖流域梅雨期所处的状态(表2).

表 2 太湖流域历年梅雨状态

Table 2 State of plum rains in Taihu Lake Basin for each year of 1954—1991

年份	状态	年份	状态	年份	状态	年份	状态	年份	状态	年份	状态	年份	状态	年份	状态
1954	4	1958	1	1962	3	1966	4	1970	4	1974	4	1978	1	1982	3
1955	3	1959	2	1963	3	1967	3	1971	4	1975	4	1979	4	1983	4
1956	4	1960	3	1964	1	1968	3	1972	3	1976	4	1980	4	1984	3
1957	3	1961	2	1965	1	1969	3	1973	3	1977	3	1981	3	1985	3

在计算1步转移概率时,可用各状态转移出现的频数来求得频率,从而可得到转移概率矩阵.例如求 p_{11} .由表2知,处于状态1的原始样本数为4(分别是1958年,1964年,1965年,1978年),而从状态1转移到状态1的样本数为1,于是 $p_{11} = M_{11}/M_1 = 1/4$.用同样方法求出其他的转移概率,从而可得到1步转移概率矩阵

$$P_1 = \begin{bmatrix} 0.250 & 0.250 & 0 & 0.500 \\ 0 & 0 & 1 & 0 \\ 0.158 & 0.052 & 0.474 & 0.316 \\ 0 & 0 & 0.667 & 0.333 \end{bmatrix}$$

由切普曼-柯尔莫哥洛夫方程可以得到2步转移概率矩阵

$$P_2 = \begin{bmatrix} 0.063 & 0.063 & 0.584 & 0.291 \\ 0.158 & 0.052 & 0.474 & 0.316 \\ 0.114 & 0.064 & 0.487 & 0.335 \\ 0.105 & 0.035 & 0.539 & 0.321 \end{bmatrix}$$

从这个随机矩阵可以看出,此链是遍历的.

现预测 1992 年梅雨期的状态.由表 2 知,1991 年系统的状态为 4.从 1 步转移矩阵可以看出:系统从状态 4 转移到状态 3 的概率为 0.667,保持不变的概率为 0.333;1990 年系统的状态为 3,经 2 步转移到 1992,从 2 步转移概率矩阵可知,从状态 3 转移到状态 1 的概率为 0.114,转移到状态 2 的概率为 0.064,转移到状态 3 的概率为 0.487,转移到状态 4 的概率为 0.335.从 1 步和 2 步转移概率矩阵可以看出,1992 年系统处于状态 3 的概率最大.由此可知 1992 年系统的状态为 3.同样可预测出 1993 年梅雨期的状态.从 2 步转移概率矩阵可以看出,1991 年的状态 4 转移到状态 1,2,3,4 的概率分别为 0.105,0.035,0.539,0.321.由此可知 1993 年的状态为 3.由太湖流域梅雨资料可知,太湖 1992 年和 1993 年的梅雨期正处于状态 3.实际与预测一致,说明用马尔可夫链来预测梅雨期是可行的^[6].

利用马尔可夫链的 1 步转移概率矩阵,还可以计算马尔可夫链的平稳分布.由定义 3 和 1 步转移概率矩阵可得

$$\begin{cases} 0.25\pi_1 + 0.158\pi_3 = \pi_1 & 0.25\pi_1 + 0.052\pi_3 = \pi_2 \\ \pi_2 + 0.474\pi_3 + 0.667\pi_4 = \pi_3 & 0.5\pi_1 + 0.316\pi_3 + 0.333\pi_4 = \pi_4 \end{cases}$$

满足 $\pi_i > 0 (i = 1, 2, 3, 4)$, $\sum_{i=1}^4 \pi_i = 1$ 的唯一解是: $\pi_1 = 0.108$, $\pi_2 = 0.054$, $\pi_3 = 0.5135$, $\pi_4 = 0.3245$.这说明系统在各状态的稳定概率分别为 0.108, 0.054, 0.5135, 0.3145.由此可以预测,梅雨期状态 1 约 9a 出现 1 次,状态 2 约 19a 出现 1 次,状态 3 约 2a 出现 1 次,状态 4 约 3a 出现 1 次.从历史资料中也可以看出,状态 3 和状态 4 出现的频率要比状态 1 和状态 2 出现的频率高很多^[7-9].

4 结 语

- a. 马尔可夫链预测是建立在对历史数据的分析统计之上的,在太湖流域气候条件稳定的情况下,历史数据越多越精确,预测也就越可靠.
- b. 梅雨状态的马尔可夫链预测是一种概率预报,其准确性和可靠性要通过大量的预测检验才能显示出来.

参考文献:

[1] 吴浩云,管唯庆.1991 年太湖流域洪水[M].北京:中国水利水电出版社,1999:34-154.
[2] 魏宗舒.概率论与数理统计教程[M].北京:高等教育出版社,1983:194-198.
[3] 刘次华.随机过程[M].2 版.武汉:华中科技大学出版社,2001:56-84.
[4] 夏乐天,朱永忠.工程随机过程[M].南京:河海大学出版社,2000:40-86.
[5] 冯利华,王德华.金衢盆地洪水的马尔可夫链预测[J].地域研究与开发,1997(2):87-89.
[6] 赵利民,周西利.马尔可夫链在大白菜年景预报中的应用[J].西北农业学报,2003(12):139-142.
[7] 王佐仁.马氏链为齐次的统计检验方法[J].统计与信息论坛,2004(5):14-15.
[8] 宫锡鸿,姜海萍.用马尔柯夫分析法预测麦蜘蛛发生趋势[J].数理统计与管理,2003(2):15-17.
[9] 咸美新.马氏链在生产决策中的应用[J].南京师大学报:自然科学版,2000(4):31-34.

Application of Markov chain to forecasting plum rain period
of Taihu Lake Basin

ZHANG Zong-guo

(College of Sciences , Hohai University , Nanjing 210098 , China)

Abstract :According to the data of beginning and ending of the plum rain period and the main periods of plum rains from 1954 to 1991 , the Markov chain was applied to forecasting the plum rain period of 1992 and 1993 for the Taihu Lake Basin , and the probability of the plum rain period in different states was analyzed based on its stability distribution . The analysis shows that , with long and accurate historical data series , the present model can give satisfactory plum rain period forecasting for the Taihu Lake Basin under the stable climate condition . However , as a probability forecasting model , the accuracy and reliability of the present model should be verified through a great amount of data forecasting and testing .

Key words :Taihu Lake Basin ; plum rain period forecasting ; Markov chain ; random matrix ; transition probability