

权马尔科夫链在北京市年降水量预测的应用

□ 徐晓宇

(中央民族大学理学院 北京 100081)

摘要: 根据降水量是一列相依随机变量的特点,利用北京市 1951 年-2007 年共 57 年的年平均降水量和其标准差,对北京市 1957 年-2007 年的年降水量划分成丰水年,偏丰水年,平水年,偏枯水年和枯水年五个等级。以规范化的各阶自相关系数为权,建立北京市降水量的权马尔科夫链预测模型。估计 2008 年北京市降水量的等级,预测结果较为满意。

关键词: 马尔科夫链 加权 年降水量 预测

中图分类号: P46

文献标识码: A

文章编号: 1007-3973 (2010) 01-111-02

1 概述

北京市位于华北平原北部,西临黄土高原、北邻内蒙古高原、东距渤海 150 公里。北京是我国的首都,我国的政治、文化的中心,也是全国交通运输的枢纽。北京地处中纬度暖温带,是典型的暖温带半湿润大陆性季风气候。气候主要特点是四季分明。夏季炎热多雨,冬季寒冷少雪,春、秋短促。降雨多集中在夏季,预测北京地区的降水量,对北京市环境保护与经济建设有一定的指导意义。

影响一个地区年降水量的因素众多,截至目前,人们运用蒙特卡洛模拟,频谱分析等方法对短期的降水概率预测取得了比较好的效果。但对于长期的年降水量而言,人们还很难通过某种方法确定出未来某一时段降水量的准确值。但是,在大多数的情况下,年降水预测只需给出在具体的某一区间段即可。

2 权马尔科夫链思想及其预测模型

2.1 马尔科夫链

马尔科夫链是一类特殊的随机过程,它最重要的特征就是“无后效性”,也就是要确定将来的状态,只需要知道它现在的状态就足够了,不需要对它以前的状态有所认识。用数学定义表示。

随机过程 $\{X_n, n=0,1,2,\dots\}$ 称为马尔科夫链,若它只取有限个或可列个值 E_0, E_1, E_2, \dots (我们以 $\{0,1,2,\dots\}$ 来标记 E_0, E_1, E_2, \dots , 并称它们是过程的状态, $\{0,1,2,\dots\}$ 或者其子集记为 S , 称为过程的状态空间), 对任意的 $n \geq 0$ 及状态 $i, j, i_0, i_1, \dots, i_{n-1}$, 有:

$$P\{X_{n+1}=j|X_0=i_0, X_1=i_1, X_2=i_2, \dots, X_n=i_n\} = P\{X_{n+1}=j|X_n=i\}$$

则称这一随机过程具有马尔科夫性,这一随机过程就称为马尔科夫链^①。

2.2 权马尔科夫链的基本思想

由马尔科夫链的定义我们可以知道,要确定将来的状态,只需知道现在的状态就可以了。但是我们可以想到,与未来状态相关的应该不只有现在的一个状态,所以我们这里引

进了自相关系数。我们用各阶自相关系数刻画了五个滞时的年降水量的相关关系及其相关的强弱。本文中,我们可以先用前面 5 年的年降雨量作为初始状态,然后按照各滞时与所求时段相依关系的强弱加权求和,就达到了充分利用已知信息进行预测的效果。

3 权马尔科夫链预测模型的方法和步骤

3.1 计算年降水量误差序列的各阶自相关系数

$$r_k = \frac{\sum_{t=1}^{n-k} (X_t - \bar{x})(X_{t+k} - \bar{x})}{\sum_{t=1}^{n-k} (X_t - \bar{x})^2}$$

r_k 表示第 k 阶自相关系数, X_t 表示第 t 年的年降水量, \bar{x} 表示近 57 年北京市年降水量的平均值, n 为年降水量序列的长度。

3.2 规范化各阶自相关系数,即:

$$w_k = |r_k| / \sum_{k=1}^m |r_k| (m \leq 5)$$

w_k 为各种滞时的马尔科夫链的权。

3.3 划分状态

建立年降水量的分级标准,根据研究问题序列的长短和具体的情况划分为几个状态空间。本文中将年降水量划分为丰水年,偏丰水年,平水年,偏枯水年和枯水年 5 种状态(对应的划分成 5 种状态空间)

3.4 确定状态

根据上面划分的 5 种状态,确定 1951 年-2007 年这 57 年的每一年的降水量所处的状态。

3.5 计算转移概率

根据上一步所得到的结果进行统计,得到不同步长的马尔科夫链的状态矩阵。

3.6 预测

以前面的 m 个时段的状态为初始状态,转移步数为 1, 2, \dots, m 。结合相应的状态概率矩阵所对应的行向量,组成一个新的概率矩阵。

将同一个状态的 m 项预测概率与所对应的加权相乘之后求和,即:

$$P_i = \sum_{k=1}^m w_k P_i^{(k)}$$

$\max\{P_i, i \text{ 为状态空间}\}$ 所对应的 i 就是对应预测时段的年降水量预测状态。

4 北京市年降水量的权马尔科夫链预测

根据北京市 1951 年-2007 年的年降水量序列, 预测 2008 年的年降水量, 由此来检验预测模型的实用性与精确性。具体资料见下表:

年份	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965
年降水量	481.5	557	657.6	960.9	931.4	1115.2	486.8	691.4	1404.6	526.4	599.5	366.9	775.5	817.2	261.4
状态	2	3	3	5	5	5	3	3	5	3	3	1	4	5	1

年份	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980
年降水量	526.7	592.4	386.5	912.9	597	511	374	698.1	474.5	391.2	682.8	778.9	664.3	718.2	380.7
状态	3	3	2	5	3	3	2	4	2	2	3	4	3	4	2

年份	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
年降水量	393.	544.	489.	488.	721	665.	683.	673.	442.	697.	747.	541.	506.	813.	572.
状态	2	4	9	8	721	3	9	3	2	3	9	5	7	2	5
状态	2	3	3	3	4	3	3	3	2	4	4	3	3	5	3

年份	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
年降水量	700.9	430.9	731.7	266.9	371.1	338.9	370.4	444.9	483.5	410.7	318	483.9
状态	4	2	4	1	1	1	1	2	2	2	1	3

由 1951 年-2007 年的年降水量序列可知, 这 57 年的年平均降水量为 590.45mm, 无偏估计的标准差 $S=213.2$, 计算误差项的各阶自相关系数以及规范化的各阶自相关系数, 如下表所示:

k	1	2	3	4	5
r_k	0.177	0.093	0.181	0.158	0.17
w_k	0.227	0.119	0.233	0.203	0.218

按照北京市年降水量的特点, 我们将序列划分为 5 个级别, 相应的马尔科夫链就是 5 个状态:

级别	状态	分级标准	数值区间 (mm)
丰水年	1	$x \geq \bar{x} + s$	$x \geq 803.65$
偏丰水年	2	$\bar{x} + 0.5s \leq x < \bar{x} + s$	$697.05 \leq x < 803.65$
平水年	3	$\bar{x} - 0.5s \leq x < \bar{x} + 0.5s$	$483.85 \leq x < 697.05$
偏枯水年	4	$\bar{x} - s \leq x < \bar{x} - 0.5s$	$377.25 \leq x < 483.85$
枯水年	5	$x < \bar{x} - s$	$x < 377.25$

由各年所处的马尔科夫链的状态, 我们可以得到以下的概率转移矩阵:

$$P^{(1)} = \begin{bmatrix} 3/8 & 1/8 & 1/4 & 1/4 & 0 \\ 1/11 & 4/11 & 3/11 & 2/11 & 1/11 \\ 2/21 & 2/21 & 10/21 & 4/21 & 1/7 \\ 1/9 & 1/3 & 1/3 & 1/9 & 1/9 \\ 1/7 & 0 & 4/7 & 0 & 2/7 \end{bmatrix} \quad P^{(2)} = \begin{bmatrix} 2/7 & 3/7 & 1/7 & 0 & 1/7 \\ 2/11 & 1/11 & 6/11 & 2/11 & 0 \\ 2/21 & 4/21 & 2/7 & 4/21 & 5/21 \\ 2/9 & 2/9 & 1/3 & 2/9 & 0 \\ 0 & 0 & 5/7 & 1/7 & 1/7 \end{bmatrix}$$

$$P^{(3)} = \begin{bmatrix} 2/7 & 5/7 & 0 & 0 & 0 \\ 1/5 & 0 & 3/5 & 1/10 & 1/10 \\ 0 & 1/7 & 2/7 & 8/21 & 4/21 \\ 2/9 & 1/9 & 5/9 & 0 & 1/9 \\ 2/7 & 1/7 & 3/7 & 0 & 1/7 \end{bmatrix} \quad P^{(4)} = \begin{bmatrix} 1/7 & 3/7 & 2/7 & 0 & 1/7 \\ 2/9 & 0 & 4/9 & 2/9 & 1/9 \\ 1/7 & 4/21 & 3/7 & 1/7 & 2/21 \\ 2/9 & 2/9 & 1/3 & 1/9 & 1/9 \\ 0 & 1/7 & 2/7 & 3/7 & 1/7 \end{bmatrix}$$

$$P^{(5)} = \begin{bmatrix} 1/7 & 2/7 & 3/7 & 1/7 & 0 \\ 1/8 & 1/8 & 1/8 & 3/8 & 1/4 \\ 1/7 & 4/21 & 10/21 & 4/21 & 0 \\ 1/9 & 2/9 & 4/9 & 2/9 & 0 \\ 1/7 & 1/7 & 2/7 & 0 & 3/7 \end{bmatrix}$$

根据 2003 年-2007 年的年降水量及其相对应的状态转移矩阵, 对北京市 2008 年的年降水量进行预测, 过程如下表所示:

初始年	状态	滞时	状态转移概率权	1	2	3	4	5
2007 年	3	1	0.227	2/21	2/21	10/21	4/21	3/21
2006 年	1	2	0.119	2/7	3/7	1/7	0	1/7
2005 年	2	3	0.233	2/10	0	6/10	1/10	1/10
2004 年	2	4	0.203	2/9	0	4/9	2/9	1/9
2003 年	2	5	0.218	1/8	1/8	1/8	3/8	2/8
加权 P_i				0.1745	0.0999	0.3824	0.1734	0.15

由上表可知 $\max\{P_i\}=0.3824$, 对应的状态的 3, 即 2008 年北京市年降水量的预测状态为平水年, 即年降水量值在 (483.85mm, 697.05mm) 之间。实际上, 2008 年北京市年降水量为 626.3mm, 与预测的情况吻合。

5 结论

本文提出的权马尔科夫链模型, 用不同步长的马尔科夫链的权之和来预测未来的状态, 既可以充分的利用了送给的已知信息, 有可以提高原有的马尔科夫链方法的精度以及合理性。利用权马尔科夫链得到的预测结果是一个区间, 相对短期预测的具体数值而言, 这种长期预测用区间作为预测范围可以提高预测的可靠性。

参考文献:

- [1] 张波, 张景肖. 应用随机过程[M]. 清华大学出版社, 2007, 3.
- [2] 苏良军. 高等数理统计[M]. 北京大学出版社, 2007, 9.
- [3] 冯耀龙, 韩文秀. 权马尔科夫链在河流丰枯状况预测中的应用[J]. 系统工程理论与实践, 1999, 10.
- [4] 汪跃军. 蚌埠市年降水量的权马尔科夫链预测[J]. 科技论坛, 2008. 1.
- [5] 李静. 时序误差权马尔科夫链在粮食丰欠状况预测中的应用[J]. 广西农业生物科学, 2001, 12.
- [6] 刘北林, 黄晓娟. 粮食安全风险动态预警模型的构建[J]. 商业经济, 2008, 10.
- [7] 刘德地, 陈晓宏. 一种北江流域年降水量的权马尔科夫链预测模型[J]. 水文, 2006, 12.