

农作物产量预报的马尔柯夫链方法

赵四强

(气象科学研究院农业气象研究所)

一、前言

马尔柯夫过程是统计学中一类重要的随机过程,时间离散、状态离散的马尔柯夫过程称为马尔柯夫链,它被广泛应用于晴雨、干旱等灾害天气的出现和持续以及天气形势转变的预报。

众所周知,在农作物产量中隐含着由于天气气候的随机变化而产生的产量波动,我们用气象产量指数 I_w 表示。 I_w 的时间序列可以看作是时间离散、状态数目有限的马尔柯夫过程,其状态之间的转变可以用马尔柯夫链转移概率计算并进行预报。本文以全国早稻产量为例,给出利用马尔柯夫链转移概率预报农作物产量丰歉的方法。

二、一阶转移概率预报法

为了得到农作物产量的马尔柯夫链,首先要对实际产量(Y)的原始时间序列进行处理,用文献[3,4]的方法均可得到农作物产量随社会生产力变化的趋势产量(Y_t),而 $I_w = Y/Y_t$ 的时间序列就是历年天气气候对产量的影响。 $I_w > 1$ 表示天气气候条件适宜, $Y > Y_t$; $I_w < 1$ 表示天气气候条件不适宜, $Y < Y_t$ 。由 I_w 的时间序列我们可以看到产量的高低也象干旱、高温或低温等天气现象一样交替或连续出现。如果将 I_w 的时间序列分成若干等级(i (或 j) = 1, 2, ...), I_w 就变成了时间离散、状态离散的马尔柯夫过程,用 I_w 表示。 I_w 由等级 i 一步转移到等级 j 的转移概率(P_{ij})由公式

$$P_{ij} = m_{ij}/m_i \quad (1)$$

计算。式中 m_i 为等级 i 出现的次数, m_{ij} 为等级 i 一步转移到等级 j 的次数。若起始状态和终止状态都有 n 个等级,则由 i 转移到 j 的一步(一阶)转移概率矩阵是

$$[P_{ij}^{(1)}] = \begin{bmatrix} P_{11}^{(1)} & P_{12}^{(1)} \dots P_{1n}^{(1)} \\ P_{21}^{(1)} & P_{22}^{(1)} \dots P_{2n}^{(1)} \\ \vdots & \vdots \\ P_{n1}^{(1)} & P_{n2}^{(1)} \dots P_{nn}^{(1)} \end{bmatrix}$$

表1给出了1963—1984年全国早稻单产(Y ,公斤/亩)、二次多项式计算的 Y_t ($Y_t = 176.594 + 5.697(t - 1962) + 0.067(t - 1962)^2$)和全国早稻单产的 I_w ,以及将 I_w 划分为5个等级得到的 I_w 。划分 I_w 1—5级的标准分别为:0.894—0.950, 0.951—0.984, 0.985

-1.033, 1.034-1.061, 1.062-1.074。

表1 全国早稻的 Y, Y_t, I_w, I_w' (1963-1984)

年	1963	1964	1965	1966	1967	1968	1969	1970	1971	1972	1973
Y	166.0	177.5	208.5	215.5	217.0	208.0	213.0	239.0	239.0	241.5	236.5
Y_t	182.6	188.5	194.5	200.6	206.8	213.2	219.7	226.3	233.1	240.0	246.9
I_w	0.909	0.942	1.072	1.074	1.049	0.976	0.970	1.056	1.025	1.006	0.958
I_w'	1	1	5	5	4	2	2	4	3	3	2
年	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984
Y	265.5	257.0	268.5	246.5	278.0	304.0	295.0	310.5	336.5	320.5	346.0
Y_t	254.1	261.3	268.7	276.2	283.9	291.6	299.5	307.6	315.7	323.9	332.3
I_w	1.045	0.983	0.999	0.894	0.979	1.042	0.985	1.010	1.066	0.989	1.041
I_w'	4	2	3	1	2	4	3	3	5	3	4

三、高阶转移概率综合预报法

根据马尔柯夫链具有无后效性的特点,预报年 t 的 I_w 等级只与前一步 $t-1$ 的等级有关,而与 $t-2, t-3, \dots$ 时刻的等级无关。为了从已知的时间序列中获得更多的信息,增加预报的依据,我们计算了高阶转移概率。如果取步长为 n ,由马尔柯夫链一步转移概率矩阵 $[P_{ij}]$ 的 n 次自乘,即

$$[P_{ij}(n)] = [P_{ij}]^n \quad (2)$$

得到 n 阶转移概率矩阵。有了高阶转移概率矩阵就可以知道 $t-2, t-3, \dots$ 时刻的每个等级(起始状态)转移到 t 时刻各个等级(终止状态)的转移概率。如果综合考虑 $t-1, t-2, \dots$ 时刻的状态对 t 时刻状态的影响,可以对 $t-1, t-2, \dots$ 时刻转移到 t 时刻的每个等级的转移概率求和,其最大值相应的等级即为 t 时刻的等级。

为了预报 1985、1986 两年全国早稻产量,我们计算了 1963-1984、1963-1985 两个时段全国早稻 I_w 的 5 阶转移概率 $P^{(1)}, P^{(2)}, P^{(3)}, P^{(4)}, P^{(5)}$, 表 2 为 1963-1984 年的 5 阶转移概率。

根据表 2 预报 1985 年全国早稻产量的步骤如下:

1. 查表 1 得经 1-5 步转移到 1985 年的起始年(84、83、82、81、80)起始状态(等级)为 4、3、5、3、3。
2. 根据起始年的起始状态查表 2, $P^{(1)}$ 第 4 行(起始状态 4)经 1 步转移到 1-5 等级的概率为 0.00, 0.50, 0.50, 0.00, 0.00; $P^{(2)}$ 第 3 行(起始状态 3)经 2 步转移到 1-5 等级的概率为 0.11, 0.22, 0.28, 0.21, 0.16, 其余类推。
3. 计算从起始年经 1-5 步转移到 1985 年出现各等级的概率和,其最大值(1.71)的等级作为 I_w 的预报值。
4. 该等级(\hat{I}_w)上下限的平均值 1.009 作为预报年 \hat{I}_w , 1985 年 $\hat{Y} = \hat{I}_w \hat{Y}_t$, \hat{Y}_t 由 1963-1984 年 Y_t 的二次多项式外推得到。

表 2 1963—1984 年的 $P^{(1)}$ 、 $P^{(2)}$ 、 $P^{(3)}$ 、 $P^{(4)}$ 、 $P^{(5)}$

终止状态 起始状态		1 2 3 4 5					1 2 3 4 5					1 2 3 4 5				
1		$P^{(1)} = \begin{bmatrix} 0.33 & 0.33 & 0.00 & 0.00 & 0.33 \\ 0.00 & 0.20 & 0.20 & 0.60 & 0.00 \\ 0.16 & 0.16 & 0.33 & 0.16 & 0.16 \\ 0.00 & 0.50 & 0.50 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.33 & 0.33 & 0.33 \end{bmatrix}$	$P^{(2)} = \begin{bmatrix} 0.11 & 0.17 & 0.17 & 0.31 & 0.22 \\ 0.03 & 0.37 & 0.40 & 0.15 & 0.03 \\ 0.11 & 0.22 & 0.28 & 0.21 & 0.16 \\ 0.08 & 0.18 & 0.26 & 0.31 & 0.08 \\ 0.05 & 0.22 & 0.38 & 0.16 & 0.16 \end{bmatrix}$	$P^{(3)} = \begin{bmatrix} 0.06 & 0.25 & 0.32 & 0.21 & 0.14 \\ 0.07 & 0.23 & 0.29 & 0.30 & 0.09 \\ 0.08 & 0.23 & 0.30 & 0.23 & 0.13 \\ 0.07 & 0.30 & 0.34 & 0.18 & 0.10 \\ 0.08 & 0.21 & 0.31 & 0.25 & 0.13 \end{bmatrix}$												
2																
3																
4																
5																
1		$P^{(4)} = \begin{bmatrix} 0.07 & 0.23 & 0.31 & 0.25 & 0.12 \\ 0.07 & 0.27 & 0.32 & 0.21 & 0.10 \\ 0.07 & 0.24 & 0.31 & 0.23 & 0.12 \\ 0.08 & 0.23 & 0.29 & 0.27 & 0.11 \\ 0.07 & 0.24 & 0.31 & 0.22 & 0.12 \end{bmatrix}$	$P^{(5)} = \begin{bmatrix} 0.07 & 0.25 & 0.31 & 0.23 & 0.11 \\ 0.07 & 0.24 & 0.30 & 0.25 & 0.11 \\ 0.07 & 0.24 & 0.31 & 0.24 & 0.12 \\ 0.07 & 0.25 & 0.32 & 0.22 & 0.11 \\ 0.07 & 0.24 & 0.31 & 0.24 & 0.12 \end{bmatrix}$													
2																
3																
4																
5																

表 3 1985、1986 年 I_{10} 出现的概率与预报

终止状态 起始状态		1985					终止状态 起始状态		1986				
		1	2	3	4	5			1	2	3	4	5
1984	4	0.00	0.50	0.50	0.00	0.00	1985	3	0.16	0.16	0.33	0.16	0.16
1983	3	0.11	0.22	0.28	0.21	0.16	1984	4	0.10	0.18	0.28	0.34	0.10
1982	5	0.08	0.21	0.31	0.25	0.13	1983	3	0.08	0.21	0.32	0.23	0.14
1981	3	0.07	0.24	0.31	0.23	0.12	1982	5	0.08	0.21	0.34	0.21	0.13
1980	3	0.17	0.24	0.31	0.24	0.12	1981	3	0.08	0.21	0.33	0.23	0.12
Σ		0.33	1.41	1.71	0.93	0.53	Σ		0.50	0.97	1.60	1.17	0.65
\hat{I}_{10}		3(0.985—1.033)					\hat{I}_{10}		3(0.985—1.033)				
I_{10}		3(0.997)					I_{10}		3(0.992)				
\hat{Y}		344.0 公斤/亩					\hat{Y}		353.0 公斤/亩				
Y		339.5 公斤/亩					Y		347.0 公斤/亩				
误差(%)		1.3					误差(%)		1.7				

表 4 5 阶转移概率预报的历史拟合

年	1968	1969	1970	1971	1972	1973	1974	1975	1976
\hat{I}_{10}	3	4	3	3	3	3	4	3	4
\hat{Y}	206.5	230.5	228.5	235.0	241.0	249.0	266.5	263.5	281.5
误差(%)	0.07	8.2	4.4	1.7	0.02	5.3	0.04	2.5	4.8
年	1977	1978	1979	1980	1981	1982	1983	1984	平均
\hat{I}_{10}	3	2	4	3	3	3	3	3	
\hat{Y}	275.0	294.5	302.5	302.0	310.5	318.5	327.0	335.5	
误差(%)	11.5	5.9	0.05	2.4	0.0	5.3	2.0	3.0	3.4

表3给出了1985、1986两年 I_w 各等级出现的概率以及根据5阶转移概率综合预报的全国早稻单产,表4为历史拟合。因1963-1967年不能计算其上一年度的5阶转移概率,故历史拟合从1968年开始。

四、因子转移概率的加权预报法

选择与预报对象相关好的因子作为预报因子,将预报对象和预报因子的原始资料都处理成等级序列资料,根据马尔柯夫链转移概率计算公式(1)计算预报因子取某个等级(起始状态)条件下,预报对象出现不同等级(终止状态)的条件概率,以此作为转移概率。对预报对象拟合好的因子给予较大权重,建立有权重的转移概率预报方程。预报步骤简述如下:

1. 产量资料及处理同表1;预报因子(x_1 :5月东亚大槽强度, x_2 :1月青藏高原500百帕 $30-40^{\circ}\text{N}$ $\sum_{i=1}^{10} H_i$)划分等级的数目和划分等级标准的原则与预报对象相同(表5)。

2. 由(1)式得 x_1, x_2 各等级对应 I_w 各等级的一步转移概率矩阵 $P_1(i, j), P_2(i, j)$ 为

$$P_1(i, j) = \begin{bmatrix} 0.00 & 0.00 & 0.66 & 0.00 & 0.33 \\ 0.00 & 0.20 & 0.40 & 0.20 & 0.20 \\ 0.16 & 0.16 & 0.16 & 0.50 & 0.00 \\ 0.00 & 0.60 & 0.20 & 0.00 & 0.20 \\ 0.66 & 0.00 & 0.00 & 0.33 & 0.00 \end{bmatrix} \quad P_2(i, j) = \begin{bmatrix} 0.33 & 0.33 & 0.00 & 0.33 & 0.00 \\ 0.20 & 0.40 & 0.20 & 0.20 & 0.00 \\ 0.00 & 0.33 & 0.50 & 0.16 & 0.00 \\ 0.20 & 0.00 & 0.40 & 0.20 & 0.20 \\ 0.00 & 0.00 & 0.00 & 0.33 & 0.66 \end{bmatrix}$$

3. 比较历年 I_w 和 I_w , 对 x_1, x_2 的拟合能力进行评分。评分标准为:与实况一致记1分,差1级记0分,差2级记-1分,差3级记-2分,每个因子得分总和(C)为该因子权重系数(表5)。

4. 根据1985年因子实况, x_1 为5级($P_1(i, j)$ 第5行) I_w 为1级的概率最大; x_2 为3级($P_2(i, j)$ 第3行) I_w 为3级的概率最大。若 x_1, x_2 实况的等级(K)相同,预报年出现K级的最大概率为

$$P_{max}(K) = C_1 P_{1max}(K) + C_2 P_{2max}(K) \tag{3}$$

式中 C_1, C_2 为 x_1, x_2 的权重系数,反之则分别计算 $C_1 P_{1max}(K)$ 和 $C_2 P_{2max}(K)$, 取其大者的等级(K=3)作为1985年 I_w 的预报值(表6)。

表5 x_1, x_2 的等级和评分

年序	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	C
I_w	1	1	5	5	4	2	2	4	3	3	2	4	2	3	1	2	4	3	3	5	3	4	
x_1 等级	5	3	2	4	3	2	3	2	2	3	4	3	4	1	5	4	3	1	2	1	4	5	
P_{1max} 的 I_w	1	4	3	2	4	3	4	3	3	4	2	4	2	3	1	2	4	3	3	3	2	1	
x_1 得分	1	-2	-1	-2	1	0	-1	0	1	0	1	1	1	1	1	1	1	1	1	-1	0	-2	3
x_2 等级	4	2	5	5	5	2	1	3	3	4	3	2	3	4	1	2	4	2	3	4	3	1	
P_{2max} 的 I_w	3	2	5	5	5	2	1	3	3	3	3	2	3	3	1	2	3	2	3	3	3	1	
x_2 得分	-1	0	1	1	0	1	0	0	1	1	0	-1	0	1	1	1	0	0	1	-1	1	-2	5

表 6 1985、1986 年预报

因子等级		P_{max}	K	C	$P_{max}(K)$	\hat{I}_n
1985	x_1 5	0.66	1	3	1.98	3
	x_2 3	0.50	3	5	2.50	
1986	x_1 3	0.42	4	-10	-0.42	3
	x_2 4	0.40	3	6	2.40	

从表 6 看出 1985、1986 两年的 \hat{I}_n 均为 3 级,与高阶转移概率预报结果一致;其历史拟合平均相对误差为 2.1%,优于高阶转移概率预报。

五、结束语

本文应用马尔柯夫链转移概率试报全国早稻产量结果表明,该方法可以用于农作物产量预报,但应注意以下问题:

1. 农作物产量的原始时间序列不是马尔柯夫链,必须对原始时间序列进行处理,例如转换为 I_n 的时间序列后才能使用。

2. 如果产量的时间序列很长,还应注意 I_n 是否为均匀的马尔柯夫链,如果不是,应分别对不同时期计算转移概率,预报年处于哪一时期就用哪一时期的转移概率作预报。

3. 由于马尔柯夫链是无后效的,前后联系非常微弱,所以用高阶转移概率预报时,外推一般不超过两年,而因子转移概率的加权预报则根据预报年因子出现的实况计算 I_n 出现各等级的概率,即考虑了当年气象条件对产量的影响,只要预报年出现的因子状态包括在已划分的几种状态之中,就可以进行外推预报。

4. 阶数较高的转移概率逐渐接近气候概率,预报价值不大,所以计算转移概率的阶数不宜过高;如果预报年出现某些等级,例如 2 级和 3 级的概率和相等,则按分级顺序取 2 级作为预报年等级。

5. 用因子转移概率的加权预报时,选择因子的个数不宜太多,一般不超过样本长度的 $\frac{1}{5}$ 。为了使因子提供较多的信息,还应注意入选因子之间的独立性。

参 考 文 献

- [1] E. Parzen, Stochastic Processes, New York, 1969.
- [2] 王宗皓、李麦村,天气预报中的概率统计方法,科学出版社,1974 年。
- [3] 冯定原、夏海峰,几种产量资料处理方法的比较,农业气象预报文集,气象出版社,1983 年。
- [4] 王霞棠,产量预报方法研究——气象产量,农业气象预报文集,气象出版社,1983 年。
- [5] 曹鸿兴,局地天气预报的数据分析方法,气象出版社,1983 年。