

0-1时间序列分析及其在害虫预测中的应用

梁其伟 李天生

(中国林业科学研究院林业研究所)

关键词 二值时间序列分析; 害虫预测; 马尾松毛虫

害虫种群由非暴发状态发展为暴发, 不仅是量的变化, 而且也是质变。这是一种两类状态的问题。在我国, 因各种原因, 至今很少地方用数值表示森林害虫种群动态的时间序列 x_1, x_2, \dots, x_N 。很多地方只有暴发和非暴发的记载。这可看成由两种状态构成的 0-1 时间序列, 0-1 时间序列分析方法为这种历史信息提供了简便的分析与预测的手段。

1 0-1时间序列简介及其预测的理论依据[1,2]

事物如果呈现两种状态, 或按某一水平值可以划分为两种状态, 我们不妨将一种状态以 1 表示, 另一种状态以 0 表示。倘若这两种状态可随时间而变, 则由此得相应的一串由 0 和 1 组成的 0-1 序列。是二值时间序列的一种。

某二值时间序列 W_1, W_2, \dots, W_N 中长为 $l+1$ 的子序列为: $\{W_1, \dots, W_l, W_{l+1}\}$; $\{W_2, \dots, W_{l+1}, W_{l+2}\}$; \dots ; $\{W_{N-l}, \dots, W_{N-1}, W_N\}$, 共有 $N-l$ 个。

这些子序列的每一个的前 l 个数与 $\{W_{N-l+1}, \dots, W_N\}$ 相比较, 累计与 $\{W_{N-l+1}, \dots, W_N\}$ 相同的子序列的个数, 记为 N_l 。

这 N_l 个子序列, 它们的前 l 个数与 $\{W_{N-l+1}, \dots, W_N\}$ 相同, 将它们最后一个数为 1 的子序列个数记为 N_1 , 最后一个数为 0 的子序列的个数记为 N_0 。

根据概率论的原理, 当 N 充分大时, 以下两个条件概率可由其频率来近似:

$$P(W_{N+1} = 1 | W_{N-l} = W_{N-l}, \dots, W_N = W_N) \approx \frac{N_1}{N_l}$$

$$P(W_{N+1} = 0 | W_{N-l} = W_{N-l}, \dots, W_N = W_N) \approx \frac{N_0}{N_l}$$

据此, 可对 0-1 时间序列作一步预报:

$$\text{若 } \frac{N_1}{N_l} > \frac{N_0}{N_l} \quad \hat{W}_{N+1} = 1,$$

$$\frac{N_1}{N_l} < \frac{N_0}{N_l} \quad \hat{W}_{N+1} = 0,$$

$$\frac{N_1}{N_l} = \frac{N_0}{N_l} \quad \text{不可预报。}$$

2 在马尾松毛虫种群预测中的应用

2.1 浙江省泗安林场马尾松毛虫发生状态的0-1时间序列预报

根据马尾松毛虫发生的特点，不是马氏过程，所以 l 不取 1，且与两代前发生关系不显著，故 l 取 2。由表 1 得 45 个 $(47 - 2 = 45)$ 长度为 3 的子序列，分别为：

010; 100; 000; 001; 010; 100; 000; 001; 010; 100; 000; 000; 001; 011; 110; 100; 001; 010; 100; 000; 000; 001; 010; 100; 000; 001; 011; 110; 100; 000; 000; 000; 001; 011; 110; 101; 010; 100; 000; 000; 000; 000; 000; 001; 011。

表 1 浙江省泗安林场南片林区历年松毛虫状态

年份	1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977
虫期	—	二	—	二	—	二	—	二	—	二	—	二
状态	0	1	0	0	0	1	0	0	1	1	0	0
年份	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
虫期	—	二	—	二	—	二	—	二	—	二	—	二
状态	0	0	0	1	1	0	0	0	1	1	0	1

注：“虫期”栏中，“—”指第一代，“二”指第二代加越冬代。

现在要预报 1989 年第二虫期的状态。由于 1988 年第二虫期的状态为 1，1989 年第一虫期的状态为 1，故 1989 年第二虫期的前两个状态为“1 1”。从上述 45 个子序列中可以看出，前两位等于“1 1”的子序列个数 $N_1 = 3$ ，这三个子序列中第三个数为 1 的子序列 $N_1 = 0$ ，第三个数为 0 的子序列 $N_0 = 3$ 。

因 $N_0 > N_1$ ，故预报 1989 年第二虫期的状态 $\hat{W}_{N+1} = 0$ 即属于非暴发状态。

1989 年第二虫期的实际状况是非暴发，与预报相符。

2.2 湖南省郴州地区马尾松毛虫发生状态0-1时间序列预报

湖南省郴州地区 1954 年至 1987 年历年马尾松毛虫发生状态的 0-1 时间序列如表 2。

表 2 湖南省郴州地区历年马尾松毛虫发生状态

年份	1954~1955	1955~1956	1956~1957	1957~1958	1958~1959	1959~1960	1960~1961	1961~1962	
状态	1	0	0	1	0	0	0	0	
年份	1962~1963	1963~1964	1964~1965	1965~1966	1966~1967	1967~1968	1968~1969	1969~1970	
状态	0	0	1	0	0	0	0	0	
年份	1970~1971	1971~1972	1972~1973	1973~1974	1974~1975	1975~1976	1976~1977	1977~1978	
状态	0	0	1	0	0	0	1	0	
年份	1978~1979	1979~1980	1980~1981	1981~1982	1982~1983	1983~1984	1984~1985	1985~1986	1986~1987
状态	0	1	0	0	0	0	1	0	0

注：1954~1955 指 1954 年第一代起至 1955 年越冬代，其余类推。

取 $l = 2$ ，则长度为 3 的子序列的个数为 $3^3 - 2 = 32$ 个，分别为：

100; 001; 010; 100; 000; 000; 000; 000; 001; 010; 100; 000; 000; 000; 000;
000; 001; 010; 100; 000; 001; 010; 100; 001; 010; 100; 000; 000; 000; 010; 100;
100。

现在预报 1987~1988 年的马尾松毛虫发生状态。因为 1985~1986 年的状态为 0，1986~1987 年的状态为 0，所以 1987~1988 年的前两个状态为“0 0”。在上述 32 个子序列中可知，前两位数等于“0 0”的子序列个数 $N_1 = 18$ 。在这 18 个子序列中，第三个数为 1 的子序列数 $N_1 = 5$ ，第三个数为 0 的子序列数 $N_0 = 13$ 。

由于 $N_0 > N_1$ ，故预报 1987~1988 年马尾松毛虫的发生状态为 0，即非暴发状态。预报结果与实际情况相符。

3 讨 论

0-1 时间序列为二值时间序列的一种。即使是数值表示的序列，也可以按某种水平值将其划分为二值时间序列，这样大有好处。可以避免原始数据量估计不准造成的技术性误差，提高分析结果的可靠性。

样本长度不够大的时候，条件概率不宜用频率来近似。但是，依据频率做判断总比没有依据要强。从数学角度来说，若样本长度为 N ， l 值可按 $N - l \geq 2^{l+1}$ 来取，若实际样本 N 不太大， l 可放宽为 $N - l - 3 \geq 2^l$ 。然而， l 到底取哪个值，还要结合害虫的发生特点，如本文的应用部分 l 取 2。0-1 时间序列分析的应用效果随着样本的长度增加而改善，样本长度越大，预测效果越佳。0-1 时间序列预测是简便易行的方法。

参 考 文 献

[1] Kedem Benjamin, 1980, Binary time series, New York Marcel Dekker, 104.

[2] 项静恬等, 1986, 动态数据处理, 时间序列分析, 气象出版社。

The Method of 0-1 Time Series Analysis and Its Application in the Prediction of Population Trend for Dendrolimus punctatus

Liang Qiwei Li Tiansheng

(The Research Institute of Forestry CAF)

Abstract The 0-1 time series analysis is a kind of binary time series analysis. It is a simple and useful mathematical method. The method was applied in the prediction of population trend of *Dendrolimus punctatus* Walker.

Key words binary time series analysis; prediction; *Dendrolimus punctatus*