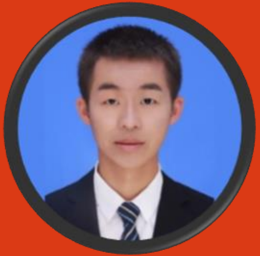
数学基础模型

时间序列模型



主讲人 张文斌

上海交通大学机械与动力工程学院博士生
曾获美国数学建模特等奖（Outstanding）
研究数学建模多年，掌握一定数模获奖技巧
熟悉数学建模方法、编程及论文写作

目录

1、时间序列模型的基本概念

时间序列是按时间顺序排列的、随时间变化且相互关联的数据序列。分析时间序列的方法构成数据分析的一个重要领域，即时间序列分析。

时间序列根据所研究的依据不同，可有不同的分类。

1．按所研究的对象的多少分，有一元时间序列和多元时间序列。

2．按时间的连续性可将时间序列分为离散时间序列和连续时间序列两种。

3．按序列的统计特性分，有平稳时间序列和非平稳时间序列。如果一个时间序列的概率分布与时间 t 无关，则称该序列为严格的（狭义的）平稳时间序列。如果序列的一、二阶矩存在，而且对任意时刻 t 满足：

（1）均值为常数

（2）协方差为时间间隔 τ 的函数。

则称该序列为宽平稳时间序列，也叫广义平稳时间序列。课程所研究的时间序列主要是宽平稳时间序列。

4．按时间序列的分布规律来分，有高斯型时间序列和非高斯型时间序列。

1、时间序列模型的基本概念

➢ 时间序列分析方法概述

时间序列预测技术就是通过对预测目标自身时间序列的处理，来研究其变化趋势的。一个时间序列往往是以下几类变化形式的叠加或耦合。

（1）长期趋势变动。它是指时间序列朝着一定的方向持续上升或下降，或停留在某一水平上的倾向，它反映了客观事物的主要变化趋势。

（2）季节变动。

（3）循环变动。通常是指周期为一年以上，由非季节因素引起的涨落起伏波形相似的波动。

（4）不规则变动。通常它分为突然变动和随机变动。

1、时间序列模型的基本概念

➢ 时间序列分析方法概述

通常用T_t表示长期趋势项，S_t表示季节变动趋势项，C_t表示循环变动趋势项，R_t表示随机干扰项。常见的时间序列模型有以下几种类型：

（1）加法模型 $y_t = T_t + S_t + C_t + R_t$

（2）乘法模型 $y_t = T_t \cdot S_t \cdot C_t \cdot R_t$

（3）混合模型 $y_t = T_t \cdot S_t + R_t$

$$y_t = S_t + T_t \cdot C_t \cdot R_t$$

其中y_t是观测目标的观测记录，

$$E(R_t) = 0, \qquad E(R_t^2) = \sigma^2$$

如果在预测时间范围以内，无突然变动且随机变动的方差σ^2较小，并且有理由认为过去和现在的演变趋势将继续发展到未来时，可用一些经验方法进行预测。

2、时间序列模型：移动平均法

移动平均法是根据时间序列资料逐渐推移，依次计算包含一定项数的时序平均数，以反映长期趋势的方法。当时间序列的数值由于受周期变动和不规则变动的影响，起伏较大，不易显示出发展趋势时，可用移动平均法，消除这些因素的影响，分析、预测序列的长期趋势。

移动平均法有简单移动平均法，加权移动平均法，趋势移动平均法等。

2、时间序列模型：移动平均法

➢ 简单移动平均法

设观测序列为$y_1, \ldots, y_T$,取移动平均的项数$N < T$。一次简单移动平均计算公式为：

$$M_t^{(1)} = \frac{1}{N}(y_t + y_{t-1} + \cdots + y_{t-N+1}) = M_{t-1}^{(1)} + \frac{1}{N}(y_t - y_{t-N})$$

当预测目标的基本趋势是在某一水平上下波动时，可用一次简单移动平均方法建立预测模型：$\hat{y}_{t+1} = M_t^{(1)} = \frac{1}{N}(y_t + y_{t-1} + \cdots + y_{t-N+1}), t = N, N+1, \ldots, T$

最近N期序列值的平均值作为未来各期的预测结果。一般N取值范围:$5 \le N \le 200$。当历史序列的基本趋势变化不大且序列中随机变动成分较多时，N 的取值应较大一些。否则N 的取值应小一些。在有确定的季节变动周期的资料中，移动平均的项数应取周期长度。选择最佳N 值的一个有效方法是，比较若干模型的预测误差。预测标准误差最小者为好。

其预测标准误差为：

$$S = \sqrt{\frac{\sum_{t=N+1}^{T}(\hat{y}_t - y_t)^2}{T - N}}$$

2、时间序列模型：移动平均法

例1-1 某企业 1 月～11 月份的销售收入时间序列如图所示。试用一次简单滑动平均法预测第 12 月份的销售收入。

月份 t	1	2	3	4	5	6
销售收入 y_t	533.8	574.6	606.9	649.8	705.1	772.0
月份 t	7	8	9	10	11	
销售收入 y_t	816.4	892.7	963.9	1015.1	1102.7	

2、时间序列模型：移动平均法

模型的建立与求解

分别去N=4,N=5的预测公式

$$\hat{y}_{t+1}^{(1)}=\frac{y_t+y_{t-1}+y_{t-2}+y_{t-3}}{4},t=4,5,\dots,11$$

$$\hat{y}_{t+1}^{(2)}=\frac{y_t+y_{t-1}+y_{t-2}+y_{t-3}+y_{t-4}}{5},t=5,\dots,11$$

简单移动平均法只适合做近期预测，而且是预测目标的发展趋势变化不大的情况。如果目标的发展趋势存在其它的变化，采用简单移动平均法就会产生较大的预测偏差和滞后。

Example_1_1

当N=4时，预测值$\hat{y}_{12}^{(1)}$ =993.9，预测的标准误差为

$$S_1=\sqrt{\frac{\sum_{t=5}^{11}(\hat{y}_t^{(1)}-y_t)^2}{11-4}}=150.5$$

当N=5时，预测值$\hat{y}_{12}^{(2)}=958.16$，预测的标准误差为

$$S_2=\sqrt{\frac{\sum_{t=6}^{11}(\hat{y}_t^{(2)}-y_t)^2}{11-5}}=182.4$$

计算结果表明， $N=4$ 时，预测的标准误差较小，所以选取 $N=4$ 。预测第 12 月份的销售收入为 993.6。

2、时间序列模型：移动平均法

➢ 加权移动平均法

在简单移动平均公式中，每期数据在求平均时的作用是等同的。但是，每期数据所包含的信息量不一样，近期数据包含着更多关于未来情况的信息。因此，把各期数据等同看待是不尽合理的，应考虑各期数据的重要性，对近期数据给予较大的权重，这就是加权移动平均法的基本思想。

设时间序列为$y_1, y_1 \dots, y_t, \dots$;加权移动平均计算公式为：

$$M_{tw} = \frac{w_1 y_1 + w_2 y_2 + \cdots + w_N y_{t-N+1}}{w_1 + w_2 + \cdots + w_N}, t \geq N$$

式中M_{tw}为t期的加权移动平均数， w_i 为y_{t-i+1} 的权数，它体现了相应的y_t 在加权平均数中的重要性。

利用加权移动平均数来做预测，其预测公式为： $\hat{y}_{t+1} = M_{tw}$

即以第t期的加权移动平均数作为第t+1期的预测值。

2、时间序列模型：移动平均法

例1-2 我国 1979～1988 年原煤产量如表所示，试用加权移动平均法预测 1989 年的产量。

年份	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988
原煤产量 y_t	6.35	6.20	6.22	6.66	7.15	7.89	8.72	8.94	9.28	9.8
三年加权移动平均预测值				6.235	6.4367	6.8317	7.4383	8.1817	8.6917	9.0733
相对误差（%）				6.38	9.98	13.41	14.7	8.48	6.34	7.41

2、时间序列模型：移动平均法

模型的建立与求解

取$w_1 = 3, w_2 = 2, w_3 = 1$，按预测公式

$$\hat{y}_{t+1} = \frac{3y_t + 2y_{t-1} + y_{t-2}}{3+2+1}$$

计算三年加权移动平均预测值，其结果列于表中。1989 年我国原煤产量的预测

$$\hat{y}_{1989} = \frac{3\times 9.8 + 2\times 9.28 + 8.94}{6} = 9.48$$

这个预测值偏低，可以修正。其方法是：先计算各年预测值与实际值的相对误差，例如 1982 年为

$$\frac{6.66-6.235}{6.66} = 6.38\%$$

将相对误差列于表中，再计算总的平均相对误差。

$$\left(1 - \frac{\sum \hat{y}_t}{\sum y_t}\right) \times 100\% = 9.5\%$$

由于总预测值的平均值比实际值低 9.5%，所以可将 1989 年的预测值修正为

$$\frac{9.48}{1-9.5\%} = 10.4788$$

3、时间序列模型：指数平滑法

一次移动平均实际上认为最近 N 期数据对未来值影响相同，都加权$1/N$；而 N 期以前的数据对未来值没有影响，加权为 0。但是，二次及更高次移动平均数的权数却不是$1/N$，且次数越高，权数的结构越复杂，但永远保持对称的权数，即两端项权数小，中间项权数大，不符合一般系统的动态性。一般说来历史数据对未来值的影响是随时间间隔的增长而递减的。所以，更切合实际的方法应是对各期观测值依时间顺序进行加权平均作为预测值。指数平滑法可满足这一要求，而且具有简单的递推形式。

指数平滑法根据平滑次数的不同，又分为一次指数平滑法、二次指数平滑法和三次指数平滑法等。

3、时间序列模型：指数平滑法

➢ 一次指数平滑法

1. 预测模型

设时间序列为$y_1, y_1 \dots, y_t, \dots$; α为加权系数，$0 < \alpha < 1$,一次指数平滑公式为：

$$S_t^{(1)} = \alpha y_t + (1-\alpha)S_{t-1}^{(1)} = S_t^{(1)} + \alpha(y_t - S_{t-1}^{(1)})$$

可由移动平均公式改进而来。

为进一步理解指数平滑的实质，把上式依次展开，有

$$S_t^{(1)} = \alpha y_t + (1-\alpha)\left[\alpha y_{t-1} + (1-\alpha)S_{t-2}^{(1)}\right] = \cdots = \alpha \sum_{j=0}^{\infty}(1-\alpha)^j y_{t-j}$$

该式表明$S_t^{(1)}$是全部历史数据的加权平均，加权系数分别为$\alpha, \alpha(1-\alpha), \alpha(1-\alpha)^2$；且$\sum_{j=0}^{\infty}(1-\alpha)^j = 1$，由于加权系数符合指数规律，又具有平滑数据的功能，故称为指数平滑。

3、时间序列模型：指数平滑法

➢ 一次指数平滑法

2. 加权系数的选择

以这种平滑值进行预测，就是一次指数平滑法。预测模型为

$\hat{y}_{t+1} = S_t^{(1)}$ 即 $\hat{y}_{t+1} = \hat{y}_t + \alpha(y_t - \hat{y}_t)$

在进行指数平滑时，加权系数的选择是很重要的。由上式可以看出，α的大小规定了在新预测值中新数据和原预测值所占的比重。α值越大，新数据所占的比重就愈大，原预测值所占的比重就愈小，反之亦然。

若选取α=0，则 $\hat{y}_{t+1} = \hat{y}_t$，即下期预测值就等于本期预测值，在预测过程中不考虑任何新信息；若选取α=1，$\hat{y}_{t+1} = y_t$，即下期预测值就等于本期观测值，完全不相信过去的信息。这两种极端情况很难做出正确的预测。因此，α 值应根据时间序列的具体性质在 0～1 之间选择。具体如何选择一般可遵循下列原则：①如果时间序列波动不大，比较平稳，则α 应取小一点，如（0.1～0.5）。以减少修正幅度，使预测模型能包含较长时间序列的信息；②如果时间序列具有迅速且明显的变动倾向，则α 应取大一点，如（0.6～0.8）。使预测模型灵敏度高一些，以便迅速跟上数据的变化。

3、时间序列模型：指数平滑法

➢ 一次指数平滑法

3．初始值的确定

用一次指数平滑法进行预测，除了选择合适的α外，还要确定初始值$S_0^{(1)}$。初始值是由预测者估计或指定的。当时间序列的数据较多，比如在 20 个以上时，初始值对以后的预测值影响很少，可选用第一期数据为初始值。如果时间序列的数据较少，在 20个以下时，初始值对以后的预测值影响很大，这时，就必须认真研究如何正确确定初始值。一般以最初几期实际值的平均值作为初始值。

3、时间序列模型：指数平滑法

➢ 一次指数平滑法

例1-3 某市 1976～1987 年某种电器销售额如表所示。试预测 1988 年该电器销售额。

年份	t	实际销售额 y_t
1976	1	50
1977	2	52
1978	3	47
1979	4	51
1980	5	49
1981	6	48
1982	7	51
1983	8	40
1984	9	48
1985	10	52
1986	11	51
1987	12	59

3、时间序列模型：指数平滑法

➢ 一次指数平滑法

例1−3

Example_1_3

从表中可以看出，α= 0.2,0.5 和0.8 时，预测值是很不相同的。究竟α取何值为好，可通过计算它们的预测标准误差 S，选取使 S 较小的那个α值。预测的标准误差见表

α	0.2	0.5	0.8
S	4.5029	4.5908	4.8426

计算结果表明：α= 0.2 时， S 较小，故选取α = 0.2， 预测 1988 年该电器销售额为 $\hat{y}_{1988}$ = 51.1754 。

4、时间序列模型：自适应滤波法

➢ **基本过程**

自适应滤波法与移动平均法、指数平滑法一样，也是以时间序列的历史观测值进行某种加权平均来预测的，它要寻找一组“最佳”的权数，其办法是先用一组给定的权数来计算一个预测值，然后计算预测误差，再根据预测误差调整权数以减少误差。这样反复进行，直至找出一组“最佳”权数，使误差减少到最低限度。由于这种调整权数的过程与通讯工程中的传输噪声过滤过程极为接近，故称为**自适应滤波法**。

自适应滤波法的基本预测公式为

$$\hat{y}_{t+1} = w_1 y_t + w_2 y_{t-1} + \cdots + w_N y_{t-N+1} = \sum_{i=1}^{N} w_i y_{t-i+1}$$

式中$\hat{y}_{t+1}$为第t+1期的预测值， w_i为t-i+1期的观测值权数， y_{t-i+1}为t-i+1期的观测值，N为权数个数。其调整权数公式为

$$w_i' = w_i + 2k e_{i+1} y_{t-i+1}$$

4、时间序列模型：自适应滤波法

例1-4 下面举一个简单的例子来说明此法的全过程。设有一个时间序列包括 10 个观测值，如下表所示。试用自适应滤波法，以两个权数来求第 11 期的预测值。

时期 t	1	2	3	4	5	6	7	8	9	10
观测值 y_t	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0

4、时间序列模型：自适应滤波法

➢ 例1-4 模型的建立与求解

本例中 $N=2$。 取初始权数 $w_1=0.5$， $w_2=0.5$， 并设 $k=0.9$。 t 的取值由 $N=2$ 开始，当$t=2$ 时：

（1）按预测公式，求第$t+1=3$期的预测值

$$\hat{y}_{t+1}=\hat{y}_3=w_1y_2+w_2y_1=0.15$$

（2）计算预测误差。

$$e_{t+1}=e_3=y_3-\hat{y}_3=0.15$$

（3）根据权数调整公式。

$$w_1'=w_1+2ke_3y_2=0.554$$

$$w_2'=w_2+2ke_3y_1=0.527$$

（1）～（3）结束，即完成了一次权数调整，然后t 进 1 再重复以前步骤。

4、时间序列模型：自适应滤波法

➢ 例1-4 模型的建立与求解

当$t = 3$时：

（1）利用所得到的权数，计算第$t+1=4$期的预测值。方法是，舍去最前面的一个观测值y_1，增加一个新的观测值y_3。即

$$\hat{y}_{t+1} = \hat{y}_4 = w_1' y_3 + w_2' y_2 = 0.2716$$

（2）计算预测误差。

$$e_{t+1} = e_4 = y_4 - \hat{y}_4 = 0.13$$

（3）根据权数调整公式。

$$w_1' = 0.554 + 2 \times 0.13 \times 0.9 \times 0.13 \times 0.3 = 0.624$$

$$w_2' = 0.527 + 2 \times 0.13 \times 0.9 \times 0.13 \times 0.2 = 0.564$$

4、时间序列模型：自适应滤波法

➢ 例1-4 模型的建立与求解

Example_1_4

这样进行到$t=10$时 $\hat{y}_{t+1}=\hat{y}_{11}=w_1'y_{10}+w_2'y_9$

但由于没有$t=11$的观测值 y_{11}，因此 $e_{t+1}=e_{11}=y_{11}-\hat{y}_{11}$ 无法计算。这时，第一轮的调整就此结束。把现有的新权数作为初始权数，重新开始$t=2$的过程。这样反复进行下去，到预测误差（指新一轮的预测总误差）没有明显改进时，就认为获得了一组“最佳”权数，能实际用来预测第 11 期的数值。本例在调整过程中，可使得误差降为零，而权数达到稳定不变，最后得到的“最佳”权数为 w_1'=2.0， w_2'=-1.0

用“最佳”权数预测第 11 期的取值

$\hat{y}_{11}=w_1'y_{10}+w_2'y_9$=1.1

在实际应用中，权数调整计算工作量可能很大，必须借助于计算机才能实现。

w =1.9999 -0.9999

练习题

1-1　下表为我国 1965～1985 年的发电总量资料，试用一次平滑指数、二次指数平滑法预测 1986 年和 1987 年的发电总量。

年份	t	发电总量 y_t
1965	1	676
1966	2	825
1967	3	774
1968	4	716
1969	5	940
1970	6	1159
1971	7	1384
1972	8	1524
1973	9	1668
1974	10	1688
1975	11	1958
1976	12	2031
1977	13	2234
1978	14	2566
1979	15	2820
1980	16	3006
1981	17	3093
1982	18	3277
1983	19	3514
1984	20	3770
1985	21	4107

感谢各位聆听!

Thanks for Listening

Q&A