

数学建模算法与应用

第15章 预测方法

预测学是一门研究预测理论、方法、评价及应用的新兴科学。综观预测的思维方式，其基本理论主要有惯性原理、类推原理和相关原理。预测的核心问题是预测的技术方法，或者说是预测的数学模型。随着经济预测、电力预测、资源预测等各种预测的兴起，预测对各种领域的重要性开始显现，预测模型也随着迅速发展。

预测的方法种类繁多，从经典的单耗法、弹性系数法、统计分析法，到目前的灰色预测法、专家系统法和模糊数学法，甚至刚刚兴起的神经网络法、优选组合法和小波分析法，据有关资料统计，预测方法多达 200 余种。因此在使用这些方法建立预测模型时，往往难以正确地判断该用哪种方法，从而不能准确地建立模型，达到要求的效果。不过预测的方法虽然很多，但各种方法多有各自的研究特点、优缺点和适用范围。

15.1 微分方程模型

当我们描述实际对象的某些特性随时间(或空间)而演变的过程、分析它的变化规律、预测它的未来性态、研究它的控制手段时，通常要建立对象的动态微分方程模型。微分方程大多是物理或几何方面的典型问题，假设条件已经给出，只需用数学符号将已知规律表示出来，即可列出方程，求解的结果就是问题的答案，答案是唯一的，但是有些问题是非物理领域的实际问题，要分析具体情况或进行类比才能给出假设条件。

做出不同的假设，就得到不同的方程。比较典型的有传染病的预测模型、经济增长预测模型、兰彻斯特（Lanchester）战争预测模型、药物在体内的分布与排除预测模型、人口的预测模型、烟雾的扩散与消失预测模型等。其基本规律随着时间的增长趋势呈指数形式，根据变量的个数建立微分方程模型。微分方程模型的建立基于相关原理的因果预测法。

该方法的优点是短、中、长期的预测都适合，既能反映内部规律，反映事物的内在关系，也能分析两个因素的相关关系，精度相应的比较高，另外对模型的改进也比较容易理解和实现。该方法的缺点是虽然反映的是内部规律，但是由于方程的建立是以局部规律的独立性假定为基础，故做中长期预测时，偏差有点大，而且微分方程的解比较难以得到。

例15.1 美日硫黄岛战役模型

J. H. Engel 用二次大战末期美日硫黄岛战役中的美军战地记录，对兰彻斯特作战模型进行了验证，发现模型结果与实际数据吻合得很好。

硫黄岛位于东京以南 660 英里的海面上，是日军的重要空军基地。美军在 1945 年 2 月开始进攻，激烈的战斗持续了一个月，双方伤亡惨重，日方守军 21500 人全部阵亡或被俘，美方投入兵力 73000 人，伤亡 20265 人，战争进行到 28 天时美军宣布占领该岛，实际战斗到 36 天才停止。美军的战地记录有按天统计的战斗减员和增援情况。日军没有后援，战地记录则全部遗失。

用 $A(t)$ 和 $J(t)$ 表示美军和日军第 t 天的人数，忽略双方的非战斗减员，则

$$\begin{cases} \frac{dA(t)}{dt} = -aJ(t) + u(t), \\ \frac{dJ(t)}{dt} = -bA(t), \\ A(0) = 0, \quad J(0) = 21500. \end{cases} \quad (15.1)$$

美军战地记录给出增援 $u(t)$ 为

$$u(t) = \begin{cases} 54000, & 0 \leq t < 1, \\ 6000, & 2 \leq t < 3, \\ 13000, & 5 \leq t < 6, \\ 0, & \text{其它.} \end{cases}$$

并可由每天伤亡人数算出 $A(t)$, $t = 1, 2, \dots, 36$ 。下面要利用这些实际数据代入 (15.1) 式, 算出 $A(t)$ 的理论值, 并与实际值比较。

利用给出的数据，对参数 a, b 进行估计。对 (15.1) 式两边积分，并用求和来近似代替积分，有

$$A(t) - A(0) = -a \sum_{\tau=1}^t J(\tau) + \sum_{\tau=1}^t u(\tau), \quad (15.2)$$

$$J(t) - J(0) = -b \sum_{\tau=1}^t A(\tau). \quad (15.3)$$

为估计 b 在 (15.3) 式中取 $t = 36$ ，因为 $J(36) = 0$ ，
且由 $A(t)$ 的实际数据可得 $\sum_{t=1}^{36} A(t) = 2037000$ ，于是从
(15.3) 式估计出 $b = 0.0106$ 。再把这个值代入 (15.3)
式即可算出 $J(t)$ ， $t = 1, 2, \dots, 36$ 。

然后从 (15.2) 式估计 a 。令 $t = 36$ ，得

$$a = \frac{\sum_{\tau=1}^{36} u(\tau) - A(36)}{\sum_{\tau=1}^{36} J(\tau)}, \quad (15.4)$$

其中分子是美军的总伤亡人数，为 20265 人，分母可由已经算出的 $J(t)$ 得到，为 372500 人，于是从 (15.4) 式有 $a = 0.0544$ 。

把这个值代入 (15.2) 式得

$$A(t) = -0.0544 \sum_{\tau=1}^t J(\tau) + \sum_{\tau=1}^t u(\tau) \quad (15.5)$$

由 (15.5) 式就能够算出美军人数 $A(t)$ 的理论值，与实际数据吻合得很好。

为了估计日军的人数，可以根据 (15.3) 式给出。当然也可以求微分方程组 (15.1) 的数值解，估计日军的人数。下面画出美军人数、日军人数的按时间变化曲线和微分方程组的轨线。(程序略)

15.2 灰色预测模型

灰色预测的主要特点是模型使用的不是原始数据序列，而是生成的数据序列。其核心体系是灰色模型（Grey Model，简称 GM），即对原始数据作累加生成（或其它方法生成）得到近似的指数规律再进行建模的方法。

优点是不需要很多的数据，一般只需要 4 个数据就够，能解决历史数据少、序列的完整性及可靠性低的问题；能利用微分方程来充分挖掘系统的本质，精度高；能将无规律的原始数据进行生成得到规律性较强的生成序列，运算简便，易于检验，具有不考虑分布规律，不考虑变化趋势。缺点是只适用于中短期的预测，只适合指数增长的预测。

15.2.1 GM(1,1)预测模型

GM(1,1)表示模型是 1 阶微分方程，且只含 1 个变量的灰色模型。

15.2.1.1 GM(1,1)模型预测方法

定义 15.1 已知参考数据列

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)),$$

1 次累加生成序列 (1—AGO)

$$\begin{aligned} x^{(1)} &= (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)) \\ &= (x^{(0)}(1), x^{(0)}(1) + x^{(0)}(2), \dots, x^{(0)}(1) + \dots + x^{(0)}(n)), \end{aligned}$$

其中 $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i)$ ($k = 1, 2, \dots, n$)。 $x^{(1)}$ 的均值生成序列

$$z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n)),$$

其中 $z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1)$, $k = 2, 3, \dots, n$ 。

建立灰微分方程

$$x^{(0)}(k) + az^{(1)}(k) = b, \quad k = 2, 3, \dots, n,$$

相应的白化微分方程为

$$\frac{dx^{(1)}}{dt} + ax^{(1)}(t) = b. \quad (15.6)$$

记 $u = [a, b]^T$, $Y = [x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)]^T$,

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}, \quad \text{则由最小二乘法, 求得使}$$

$J(u) = (Y - Bu)^T (Y - Bu)$ 达到最小值的 u 的估计值

$$\hat{u} = [\hat{a}, \hat{b}]^T = (B^T B)^{-1} B^T Y.$$

建立灰微分方程

$$x^{(0)}(k) + az^{(1)}(k) = b, \quad k = 2, 3, \dots, n,$$

相应的白化微分方程为

$$\frac{dx^{(1)}}{dt} + ax^{(1)}(t) = b. \quad (15.6)$$

记 $u = [a, b]^T$, $Y = [x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)]^T$,

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}, \quad \text{则由最小二乘法, 求得使}$$

$J(u) = (Y - Bu)^T (Y - Bu)$ 达到最小值的 u 的估计值

$$\hat{u} = [\hat{a}, \hat{b}]^T = (B^T B)^{-1} B^T Y.$$

于是求解方程 (15.6) 得

$$\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{\hat{b}}{\hat{a}} \right) e^{-\hat{a}k} + \frac{\hat{b}}{\hat{a}}, \quad k = 0, 1, \dots, n-1, \dots.$$

15.2.1.2 GM(1,1)模型预测步骤

1. 数据的检验与处理

首先,为了保证建模方法的可行性,需要对已知数据列作必要的检验处理。设参考数据为 $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$, 计算序列的级比

$$\lambda(k) = \frac{x^{(0)}(k-1)}{x^{(0)}(k)}, \quad k = 2, 3, \dots, n.$$

如果所有的级比 $\lambda(k)$ 都落在可容覆盖 $\Theta = (e^{-\frac{2}{n+1}}, e^{\frac{2}{n+2}})$ 内，则序列 $x^{(0)}$ 可以作为模型 GM(1,1) 的数据进行灰色预测。否则，需要对序列 $x^{(0)}$ 做必要的变换处理，使其落入可容覆盖内。即取适当的常数 c ，作平移变换

$$y^{(0)}(k) = x^{(0)}(k) + c, \quad k = 1, 2, \dots, n,$$

使序列 $y^{(0)} = (y^{(0)}(1), y^{(0)}(2), \dots, y^{(0)}(n))$ 的级比

$$\lambda_y(k) = \frac{y^{(0)}(k-1)}{y^{(0)}(k)} \in \Theta, \quad k = 2, 3, \dots, n.$$

2. 建立模型

按 (15.6) 式建立 GM(1,1) 模型, 则可以得到预测值

$$\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{\hat{b}}{\hat{a}} \right) e^{-\hat{a}k} + \frac{\hat{b}}{\hat{a}}, \quad k = 0, 1, \dots, n-1, \dots,$$

而且

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k), \quad k = 1, 2, \dots, n-1, \dots。$$

3. 检验预测值

(1) 残差检验

令残差为 $\varepsilon(k)$ ，计算

$$\varepsilon(k) = \frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)}, \quad k = 1, 2, \dots, n,$$

这里 $\hat{x}^{(0)}(1) = x^{(0)}(1)$ ，如果 $\varepsilon(k) < 0.2$ ，则可认为达到一般要求；如果 $\varepsilon(k) < 0.1$ ，则认为达到较高的要求。

(2) 级比偏差值检验

首先由参考数据 $x^{(0)}(k-1)$, $x^{(0)}(k)$ 计算出级比 $\lambda(k)$, 再用发展系数 a 求出相应的级比偏差

$$\rho(k) = 1 - \left(\frac{1 - 0.5a}{1 + 0.5a} \right) \lambda(k),$$

如果 $\rho(k) < 0.2$, 则可认为达到一般要求; 如果 $\rho(k) < 0.1$, 则认为达到较高的要求。

4. 预测预报

由 GM(1,1) 模型得到指定时区内的预测值, 根据实际问题的需要, 给出相应的预测预报。

15.2.1.3 GM(1,1)模型预测实例

例 15.2 北方某城市 1986~1992 年道路交通噪声平均声级数据见表 15.1。

表 15.1 城市交通噪声数据[dB(A)]

序号	年份	L_{eq}	序号	年份	L_{eq}
1	1986	71.1	5	1990	71.4
2	1987	72.4	6	1991	72.0
3	1988	72.4	7	1992	71.6
4	1989	72.1			

1. 级比检验

建立交通噪声平均声级数据时间序列如下

$$\begin{aligned}x^{(0)} &= (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(7)) \\ &= (71.1, 72.4, 72.4, 72.1, 71.4, 72.0, 71.6).\end{aligned}$$

(1) 求级比 $\lambda(k)$

$$\begin{aligned}\lambda(k) &= \frac{x^{(0)}(k-1)}{x^{(0)}(k)}, \\ \lambda &= (\lambda(2), \lambda(3), \dots, \lambda(7)) \\ &= (0.982, 1, 1.0042, 1.0098, 0.9917, 1.0056).\end{aligned}$$

(2) 级比判断

由于所有的 $\lambda(k) \in [0.982, 1.0098]$, $k = 2, \dots, 7$,
故可以用 $x^{(0)}$ 作满意的 GM(1,1) 建模。

2. GM(1,1)建模

(1) 对原始数据 $x^{(0)}$ 作一次累加, 得到
 $x^{(1)} = (71.1, 143.5, 215.9, 288, 359.4, 431.4, 503).$

(2) 构造数据矩阵 B 及数据向量 Y

$$B = \begin{bmatrix} -\frac{1}{2}(x^{(1)}(1) + x^{(1)}(2)) & 1 \\ -\frac{1}{2}(x^{(1)}(2) + x^{(1)}(3)) & 1 \\ \vdots & \vdots \\ -\frac{1}{2}(x^{(1)}(6) + x^{(1)}(7)) & 1 \end{bmatrix}, \quad Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(7) \end{bmatrix}.$$

(3) 计算

$$\hat{u} = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = (B^T B)^{-1} B^T Y = \begin{bmatrix} 0.0023 \\ 72.6573 \end{bmatrix},$$

于是得到 $\hat{a} = 0.0023$, $\hat{b} = 72.6573$ 。

(4) 建立模型

$$\frac{dx^{(1)}}{dt} + \hat{a}x^{(1)} = \hat{b},$$

求解得

$$\begin{aligned}\hat{x}^{(1)}(k+1) &= \left(x^{(0)}(1) - \frac{\hat{b}}{\hat{a}} \right) e^{-\hat{a}k} + \frac{\hat{b}}{\hat{a}}. \\ &= -30929e^{-0.0023k} + 31000\end{aligned}\tag{15.7}$$

(5) 求生成序列预测值 $\hat{x}^{(1)}(k+1)$ 及模型还原值 $\hat{x}^{(0)}(k+1)$, 令 $k=1,2,3,4,5,6$, 由 (15.7) 式的时间响应函数可算得 $\hat{x}^{(1)}$, 其中取 $\hat{x}^{(1)}(1) = \hat{x}^{(0)}(1) = x^{(0)}(1) = 71.1$, 由

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k),$$

取 $k=1,2,3,4,5,6$, 得

$$\begin{aligned}\hat{x}^{(0)} &= (\hat{x}^{(0)}(1), \hat{x}^{(0)}(2), \dots, \hat{x}^{(0)}(7)) \\ &= (71.1, 72.4, 72.2, 72.1, 71.9, 71.7, 71.6)'\end{aligned}$$

3. 模型检验

模型的各种检验指标值的计算结果见表 15.2。

表 15.2 GM(1,1)模型检验表

序号	年份	原始值	预测值	残差	相对误差	级比偏差
1	1986	71.1	71.1	0	0	
2	1987	72.4	72.4057	-0.0057	0.01%	0.0203
3	1988	72.4	72.2362	0.1638	0.23%	0.0023
4	1989	72.1	72.0671	0.0329	0.05%	-0.0018
5	1990	71.4	71.8984	-0.4984	0.7%	-0.0074
6	1991	72.0	71.7301	0.2699	0.37%	0.0107
7	1992	71.6	71.5622	0.0378	0.05%	-0.0032

经验证，该模型的精度较高，可进行预测和预报。

(程序略)

15.2.2 GM(2,1)、DGM和Verhulst模型

GM(1,1)模型适用于具有较强指数规律的序列，只能描述单调的变化过程，对于非单调的摆动发展序列或有饱和的 S 形序列，可以考虑建立 GM(2,1)，DGM 和 Verhulst 模型。

1 GM(2,1)模型

定义 15.2 设原始序列

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)),$$

其 1 次累加生成序列 (1-AGO) $x^{(1)}$ 和 1 次累减生成序列 (1-IAGO) $\alpha^{(1)}x^{(0)}$ 分别为

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)),$$

和

$$\alpha^{(1)}x^{(0)} = (\alpha^{(1)}x^{(0)}(2), \dots, \alpha^{(1)}x^{(0)}(n)),$$

其中

$$\alpha^{(1)}x^{(0)}(k) = x^{(0)}(k) - x^{(0)}(k-1), \quad k = 2, 3, \dots, n,$$

$x^{(1)}$ 的均值生成序列为

$$z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n)),$$

则称

$$\alpha^{(1)}x^{(0)}(k) + a_1x^{(0)}(k) + a_2z^{(1)}(k) = b \quad (15.8)$$

为 GM(2,1)模型。

定义 15.3 称

$$\frac{d^2 x^{(1)}}{dt^2} + a_1 \frac{dx^{(1)}}{dt} + a_2 x^{(1)} = b \quad (15.9)$$

为 GM(2,1)模型的白化方程。

定理 15.1 设 $x^{(0)}$, $x^{(1)}$, $\alpha^{(1)}x^{(0)}$ 如定义 15.2 所述,
且

$$B = \begin{bmatrix} -x^{(0)}(2) & -z^{(1)}(2) & 1 \\ -x^{(0)}(3) & -z^{(1)}(3) & 1 \\ \vdots & \vdots & \vdots \\ -x^{(0)}(n) & -z^{(1)}(n) & 1 \end{bmatrix},$$
$$Y = \begin{bmatrix} \alpha^{(1)}x^{(0)}(2) \\ \alpha^{(1)}x^{(0)}(3) \\ \vdots \\ \alpha^{(1)}x^{(0)}(n) \end{bmatrix} = \begin{bmatrix} x^{(0)}(2) - x^{(0)}(1) \\ x^{(0)}(3) - x^{(0)}(2) \\ \vdots \\ x^{(0)}(n) - x^{(0)}(n-1) \end{bmatrix},$$

则 GM(2,1)模型参数序列 $u = [a_1, a_2, b]^T$ 的最小二乘估计为 $\hat{u} = (B^T B)^{-1} B^T Y$.

例 15.3 已知 $x^{(0)} = (41, 49, 61, 78, 96, 104)$, 试建立 GM(2,1)模型。

解 $x^{(0)}$ 的 1-AGO 序列 $x^{(1)}$ 和 1-IAGO 序列 $\alpha^{(1)}x^{(0)}$ 分别为

$$x^{(1)} = (41, 90, 151, 229, 325, 429),$$

$$\alpha^{(1)}x^{(0)} = (8, 12, 17, 18, 8),$$

$x^{(1)}$ 的均值生成序列

$$z^{(1)} = (65.5, 120.5, 190, 277, 377)$$

$$B = \begin{bmatrix} -x^{(0)}(2) & -z^{(1)}(2) & 1 \\ -x^{(0)}(3) & -z^{(1)}(3) & 1 \\ \vdots & \vdots & \vdots \\ -x^{(0)}(6) & -z^{(1)}(6) & 1 \end{bmatrix} = \begin{bmatrix} -49 & -65.5 & 1 \\ -61 & -120.5 & 1 \\ -78 & -190 & 1 \\ -96 & -277 & 1 \\ -104 & -377 & 1 \end{bmatrix},$$

$$Y = [8, 12, 17, 18, 8]^T,$$

$$\hat{u} = \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{b} \end{bmatrix} = (B^T B)^{-1} B^T Y = \begin{bmatrix} -1.0922 \\ 0.1959 \\ -31.7983 \end{bmatrix},$$

故得 GM(2,1)白化模型

$$\frac{d^2 x^{(1)}}{dt^2} - 1.0922 \frac{dx^{(1)}}{dt} + 0.1959 x^{(1)} = -31.7983.$$

利用边界条件 $x^{(1)}(1) = 41$, $x^{(1)}(6) = 429$, 解之得

$$x^{(1)}(t) = 203.85e^{0.22622t} - 0.5325e^{0.86597t} - 162.317,$$

于是 GM(2,1)时间响应式

$$\hat{x}^{(1)}(k+1) = 203.85e^{0.22622k} - 0.5325e^{0.86597k} - 162.317.$$

所以

$$\hat{x}^{(1)} = (41, 92, 155, 232, 325, 429).$$

做 IAGO 还原，有

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k),$$

$$\hat{x}^{(0)} = (41, 51, 63, 77, 92, 104).$$

计算结果见表 15.3(程序略)。

表 15.3 误差检验表

序号	实际数据 $x^{(0)}$	预测数据 $\hat{x}^{(0)}$	残差 $x^{(0)} - \hat{x}^{(0)}$	相对误差 Δ_k
2	49	51	-2	4.1%
3	61	63	-2	3.3%
4	78	77	1	1.3%
5	96	92	4	4.2%
6	104	104	0	0

2 DGM(2,1)模型

定义 15.4 设原始序列

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)),$$

其 1-AGO 序列 $x^{(1)}$ 和 1-IAGO 序列 $\alpha^{(1)}x^{(0)}$ 分别为

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)),$$

和

$$\alpha^{(1)}x^{(0)} = (\alpha^{(1)}x^{(0)}(2), \dots, \alpha^{(1)}x^{(0)}(n)),$$

则称

$$\alpha^{(1)}x^{(0)}(k) + ax^{(0)}(k) = b \quad (15.10)$$

为 DGM(2,1)模型。

定义15.5 称

$$\frac{d^2 x^{(1)}}{dt} + a \frac{dx^{(1)}}{dt} = b \quad (15.11)$$

为DGM(2,1)模型的白化方程。

定理15.2 若 $u = [a, b]^T$ 为模型中的参数序列, 而 $x^{(0)}, x^{(1)}, \alpha^{(1)} x^{(0)}$ 如定义15.4所述,

$$B = \begin{bmatrix} -x^{(0)}(2) & 1 \\ -x^{(0)}(3) & 1 \\ \vdots & \vdots \\ -x^{(0)}(n) & 1 \end{bmatrix},$$

$$Y = \begin{bmatrix} \alpha^{(1)} x^{(0)}(2) \\ \alpha^{(1)} x^{(0)}(3) \\ \vdots \\ \alpha^{(1)} x^{(0)}(n) \end{bmatrix} = \begin{bmatrix} x^{(0)}(2) - x^{(0)}(1) \\ x^{(0)}(3) - x^{(0)}(2) \\ \vdots \\ x^{(0)}(n) - x^{(0)}(n-1) \end{bmatrix},$$

则 DGM(2,1)模型 $\alpha^{(1)} x^{(0)}(k) + ax^{(0)}(k) = b$ 中参数的最小二乘估计满足 $\hat{u} = [\hat{a}, \hat{b}]^T = (B^T B)^{-1} B^T Y$.

定理 15.3 设 $x^{(0)}$ 为原始序列, $x^{(1)}$ 为 $x^{(0)}$ 的 1-AGO 序列, $\alpha^{(1)}x^{(0)}$ 为 $x^{(0)}$ 的 1-IAGO 序列, \hat{a}, \hat{b} 如定理 15.2 所述, 则

(1) 白化方程 $\frac{d^2 x^{(1)}}{dt} + \hat{a} \frac{dx^{(1)}}{dt} = \hat{b}$ 的解 (时间响应函数)

为

$$\hat{x}^{(1)}(t) = \left(\frac{\hat{b}}{\hat{a}^2} - \frac{x^{(0)}(1)}{\hat{a}} \right) e^{-\hat{a}t} + \frac{\hat{b}}{\hat{a}} t + \frac{1 + \hat{a}}{\hat{a}} x^{(0)}(1) - \frac{\hat{b}}{\hat{a}^2} \quad (15.12)$$

(2) DGM(2,1)模型 $\alpha^{(1)}x^{(0)}(k) + \hat{a}x^{(0)}(k) = \hat{b}$ 的时间响应序列为

$$\hat{x}^{(1)}(k+1) = \left(\frac{\hat{b}}{\hat{a}^2} - \frac{x^{(0)}(1)}{\hat{a}} \right) e^{-\hat{a}k} + \frac{\hat{b}}{\hat{a}}k + \frac{1+\hat{a}}{\hat{a}}x^{(0)}(1) - \frac{\hat{b}}{\hat{a}^2}.$$

(15.13)

(3) 还原值为

$$\hat{x}^{(0)}(k+1) = \alpha^{(1)}\hat{x}^{(1)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k).$$

(15.14)

例15.4 试对序列

$$x^{(0)} = (2.874, 3.278, 3.39, 3.679, 3.77, 3.8)$$

建立 DGM(2,1)模型。

解 因为

$$B = \begin{bmatrix} -3.284 & -3.39 & -3.679 & -3.77 & -3.8 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}^T,$$

$$Y = [0.404, 0.112, 0.289, 0.091, 0.03]^T,$$

$$\hat{u} = \begin{bmatrix} a \\ b \end{bmatrix} = (B^T B)^{-1} B^T Y = \begin{bmatrix} 0.424 \\ 1.7046 \end{bmatrix},$$

得DGM模型的时间响应序列为

$$\hat{x}^{(1)}(k+1) = 2.7033e^{-0.424k} + 4.0202k + 0.1707,$$

所以

$$\hat{x}^{(1)} = (2.874, 5.96, 9.3688, 12.9889, 16.7473, 20.5962),$$

作1-IAGO还原

$$\hat{x}^{(0)}(k) = \hat{x}^{(1)}(k) - \hat{x}^{(1)}(k-1),$$

得

$$\hat{x}^{(0)} = (2.874, 3.086, 3.4088, 3.6201, 3.7584, 3.8488).$$

计算结果见表15.4。

表 15.4 误差检验表

序号	原始数据 $x^{(0)}$	预测数据 $\hat{x}^{(0)}$	残差 $x^{(0)} - \hat{x}^{(0)}$	相对误差 Δ_k
2	3.278	3.086	0.192	5.9%
3	3.39	3.4088	-0.0188	0.6%
4	3.679	3.6201	0.0589	1.6%
5	3.77	3.7584	0.0116	0.3%
6	3.8	3.8488	-0.0488	1.3%

3 灰色Verhulst预测模型

Verhulst 模型主要用来描述具有饱和状态的过程，即 S 形过程，常用于人口预测、生物生长、繁殖预测及产品经济寿命预测等。

Verhulst 模型的基本原理和计算方法简介如下。

定义 15.6 设 $x^{(0)}$ 为原始数据序列

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)),$$

$x^{(1)}$ 为 $x^{(0)}$ 的一次累加生成 (1-AGO) 序列,

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)),$$

$z^{(1)}$ 为 $x^{(1)}$ 的均值生成序列,

$$z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n)).$$

则称

$$x^{(0)} + az^{(1)} = b(z^{(1)})^2 \quad (15.15)$$

为灰色 Verhulst 模型, a 和 b 为参数。称

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b(x^{(1)})^2 \quad (15.16)$$

为灰色 Verhulst 模型的白化方程, 其中 t 为时间。

定理 15.4 设灰色 Verhulst 模型如上所述, 若

$$u = [a, b]^T$$

为参数列, 且

$$B = \begin{bmatrix} -z^{(1)}(2) & (z^{(1)}(2))^2 \\ -z^{(1)}(3) & (z^{(1)}(3))^2 \\ \vdots & \vdots \\ -z^{(1)}(n) & (z^{(1)}(n))^2 \end{bmatrix}, \quad Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix},$$

则参数列 u 的最小二乘估计满足

$$\hat{u} = [\hat{a}, \hat{b}]^T = (B^T B)^{-1} B^T Y.$$

定理 15.5 设灰色 Verhulst 模型如上所述, 则白化方程的解为

$$x^{(1)}(t) = \frac{\hat{a}x^{(0)}(1)}{\hat{b}x^{(0)}(1) + (\hat{a} - \hat{b}x^{(0)}(1))e^{\hat{a}t}}, \quad (15.17)$$

灰色 Verhulst 模型的时间响应序列为

$$\hat{x}^{(1)}(k+1) = \frac{\hat{a}x^{(0)}(1)}{\hat{b}x^{(0)}(1) + (\hat{a} - \hat{b}x^{(0)}(1))e^{\hat{a}k}}, \quad (15.18)$$

累减还原式为

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k). \quad (15.19)$$

例15.5 试对序列

$$x^{(0)} = (4.93, 2.33, 3.87, 4.35, 6.63, \\ 7.15, 5.37, 6.39, 7.81, 8.35)$$

建立Verhulst模型。

表 15.5 原始数据、预测值及 Verhulst 模型误差

序号	原始数据 $x^{(0)}$	预测值 $\hat{x}^{(0)}$	残差 $x^{(0)} - \hat{x}^{(0)}$	相对误差 Δ_k
1	4.93	4.93	0	0%
2	2.33	1.952177	0.377823	16.22%
3	3.87	2.635709	1.234291	31.89%
4	4.35	3.48164	0.86836	19.96%
5	6.63	4.46864	2.16136	32.60%
6	7.15	5.528334	1.621666	22.68%
7	5.37	6.536384	-1.16638	21.72%
8	6.39	7.326754	-0.93675	14.66%
9	7.81	7.73743	0.07257	0.9%
10	8.35	7.673378	0.676622	8.10%

解 计算得 1 次累加序列

$$\mathbf{x}^{(1)} = (4.93, 7.26, 11.13, 15.48, \\ 22.11, 29.26, 34.63, 41.02, 48.83, 57.18)'$$

$\mathbf{x}^{(1)}$ 的均值生成序列

$$\mathbf{z}^{(1)} = (z^{(1)}(2), \cdots, z^{(1)}(10)) \\ = (6.095, 9.195, 13.305, 18.795, \\ 25.685, 31.945, 37.825, 44.925, 53.005)'$$

于是

$$\mathbf{B} = \begin{bmatrix} -z^{(1)}(2) & (z^{(1)}(2))^2 \\ -z^{(1)}(3) & (z^{(1)}(3))^2 \\ \vdots & \vdots \\ -z^{(1)}(10) & (z^{(1)}(10))^2 \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(10) \end{bmatrix}.$$

对参数列 $u = [a, b]^T$ 进行最小二乘估计，得

$$\hat{u} = (B^T B)^{-1} B^T Y = \begin{bmatrix} -0.3576 \\ -0.0041 \end{bmatrix}.$$

Verhulst 模型为

$$\frac{dx^{(1)}}{dt} - 0.3576x^{(1)} = -0.0041(x^{(1)})^2,$$

其时间响应为

$$\begin{aligned} \hat{x}_{k+1}^{(1)} &= \frac{\hat{a}x^{(0)}(1)}{\hat{b}x^{(0)}(1) + (\hat{a} - \hat{b}x^{(0)}(1))e^{ak}} \\ &= \frac{0.3576x^{(0)}(1)}{0.0041x^{(0)}(1) + (0.3576 - 0.0041x^{(0)}(1))e^{-0.3576k}}, \end{aligned}$$

令 $k = 0, 1, \dots, 9$ 求得 $x^{(1)}$ 的预测值

$$\hat{x}^{(1)} = (\hat{x}^{(1)}(1), \dots, \hat{x}^{(1)}(10)),$$

最后求得 $x^{(0)}$ 的预测值及误差分析数据见表 15.5 的第 3, 4, 5 列。

15.3 差分方程

15.3.1 商品销售量预测

在利用差分方程建模研究实际问题时，常常需要根据统计数据用最小二乘法来拟合出差分方程的系数。其系统稳定性讨论要用到代数方程的求根。

在利用差分方程建模研究实际问题时，常常需要根据统计数据用最小二乘法来拟合出差分方程的系数。其系统稳定性讨论要用到代数方程的求根。

例 15.6 某商品前 5 年的销售量见表 15.6。现希望根据前 5 年的统计数据预测第 6 年起该商品在各季度中的销售量。

表 15.6 前 5 年销售数据表

	第 1 年	第 2 年	第 3 年	第 4 年	第 5 年
第 1 季度	11	12	13	15	16
第 2 季度	16	18	20	24	25
第 3 季度	25	26	27	30	32
第 4 季度	12	14	15	15	17

从表 15.6 可以看出，该商品在前 5 年相同季节里的销售量呈增长趋势，而在同一年中销售量先增后减，第一季度的销售量最小而第三季度的销售量最大。预测该商品以后的销售情况，根据本例中数据的特征，可以用回归分析方法按季度建立四个经验公式，分别用来预测以后各年同一季度的销售量。

在利用差分方程建模研究实际问题时，常常需要根据统计数据用最小二乘法来拟合出差分方程的系数。其系统稳定性讨论要用到代数方程的求根。

根据 $y_t^{(1)} = 1.3t + 9.5$, 预测第六年起第一季度的销售量为 $y_6^{(1)} = 17.3$, $y_7^{(1)} = 18.6$, ...。由于数据少, 用回归分析效果不一定好。

如认为销售量并非逐年等量增长而是按前一年或前几年同期销售量的一定比例增长的，则可建立相应的差分方程模型。仍以第一季度为例，为简单起见不再引入上标，以 y_t 表示第 t 年第一季度的销售量，建立形式如下的差分公式

$$y_t = a_1 y_{t-1} + a_2,$$

或

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + a_3$$

等等。

上述差分方程中的系数不一定能使所有统计数据吻合，较为合理的办法是用最小二乘法求一组总体吻合较好的数据。以建立二阶差分方程 $y_t = a_1 y_{t-1} + a_2 y_{t-2} + a_3$ 为例，选取 a_1, a_2, a_3 使得对于已知观测数据 y_t ($t = 1, 2, 3, 4, 5$)，使

$$\sum_{t=3}^5 [y_t - (a_1 y_{t-1} + a_2 y_{t-2} + a_3)]^2$$

达到最小。

编写 Matlab 程序如下

```
y0=[11 12 13 15 16]';
```

```
y=y0(3:5);x=[y0(2:4),y0(1:3),ones(3,1)];
```

```
z=x\y
```

求得 $a_1 = z(1) = -1$, $a_2 = z(2) = 3$, $a_3 = z(3) = -8$ 。即

所求二阶差分方程为

$$y_t = -y_{t-1} + 3y_{t-2} - 8.$$

虽然这一差分方程恰好使所有统计数据吻合，但这只是一个巧合。根据这一方程，可迭代求出以后各年第一季度销售量的预测值 $y_6 = 21$, $y_7 = 19$, ...。

上述为预测各年第一季度销售量而建立的二阶差分方程，虽然其系数与前5年第一季度的统计数据完全吻合，但用于预测时预测值与事实不符。凭直觉，第六年估计值明显偏高，第七年销售量预测值甚至小于第六年。稍作分析，不难看出，如分别对每一季度建立一个差分方程，则根据统计数据拟合出的系数可能会相差甚大，但对同一种商品，这种差异应当是微小的，故应根据统计数据建立一个共用于各个季度的差分方程。

为此，将季度编号为 $t = 1, 2, \dots, 20$ ，令 $y_t = a_1 y_{t-4} + a_2$ 或 $y_t = a_1 y_{t-4} + a_2 y_{t-8} + a_3$ 等，利用全体数据来拟合，求拟合得最好的系数。以二阶差分方程为例，求 a_1, a_2, a_3 使得

$$Q(a_1, a_2, a_3) = \sum_{t=9}^{20} [y_t - (a_1 y_{t-4} + a_2 y_{t-8} + a_3)]^2$$

达到最小，计算得

$$a_1 = z(1) = 0.8737, a_2 = z(2) = 0.1941, a_3 = z(3) = 0.6957,$$

故求得二阶差分方程

$$y_t = 0.8737 y_{t-4} + 0.1941 y_{t-8} + 0.6957,$$

根据此式迭代，可求得第六年和第七年第一季度销售量的预测值为

$$y_{21} = 17.5869, \quad y_{25} = 19.1676.$$

还是较为可信的。(程序略)

15.3.2 养老保险

例 15.7 某保险公司的一份材料指出，在每月交费 200 元至 59 岁年底，60 岁开始领取养老金的约定下，男子若 25 岁起投保，届时月养老金 2282 元；假定人的寿命为 75 岁，试求出保险公司为了兑现保险责任，每月至少应有多少投资收益率？

解 设 r 表示保险金的投资收益率，缴费期间月缴费额为 p 元，领养老金期间月领取额为 q 元，缴费的月数为 N ，到75岁时领取养老金的月数为 M ，投保人在投保后第 k 个月所交保险费及收益的累计总额为 F_k ，那么容易得到数学模型为分段表示的差分方程

$$F_{k+1} = F_k(1+r) + p, \quad k = 0, 1, \dots, N-1,$$

$$F_{k+1} = F_k(1+r) - q, \quad k = N, N+1, \dots, M-1$$

这里 $p = 200$, $q = 2282$, $N = 420$, $M = 600$ 。

可推出差分方程的解 (这里 $F_0 = F_M = 0$)

$$F_k = [(1+r)^k - 1] \frac{p}{r}, \quad k = 0, 1, 2, \dots, N, \quad (15.20)$$

$$F_k = \frac{q}{r} [1 - (1+r)^{k-M}], \quad k = N+1, \dots, M, \quad (15.21)$$

由式 (15.20) 和式 (15.21) 得

$$F_N = [(1+r)^N - 1] \frac{p}{r},$$

$$F_{N+1} = \frac{q}{r} [1 - (1+r)^{N+1-M}],$$

由于 $F_{N+1} = F_N(1+r) - q$ ，可以得到如下的方程

$$\frac{q}{r}[1 - (1+r)^{N+1-M}] = [(1+r)^N - 1]\frac{p}{r}(1+r) - q,$$

化简得

$$(1+r)^M - (1 + \frac{q}{p})(1+r)^{M-N} + \frac{q}{p} = 0,$$

记 $x = 1+r$ ，代入数据得（程序略）

$$x^{600} - 12.41x^{180} + 11.41 = 0.$$

15.4 马尔科夫预测

15.4.1 马尔可夫链的定义

现实世界中有很多这样的现象，某一系统在已知现在情况的条件下，系统未来时刻的情况只与现在有关，而与过去的历史无直接关系。比如，研究一个商店的累计销售额，如果现在时刻的累计销售额已知，则未来某一时刻的累计销售额与现在时刻以前的任一时刻累计销售额无关。描述这类随机现象的数学模型称为马尔科夫模型，简称马氏模型。

定义 15.7 设 $\{\xi_n, n = 1, 2, \dots\}$ 是一个随机序列, 状态空间 E 为有限或可列集, 对于任意的正整数 m, n , 若 $i, j, i_k \in E (k = 1, \dots, n-1)$, 有

$$P\{\xi_{n+m} = j \mid \xi_n = i, \xi_{n-1} = i_{n-1}, \dots, \xi_1 = i_1\} = P\{\xi_{n+m} = j \mid \xi_n = i\} \quad (15.22)$$

则称 $\{\xi_n, n = 1, 2, \dots\}$ 为一个马尔可夫链 (简称马氏链),
(15.22) 式称为马氏性。

事实上，可以证明若等式 (15.22) 对于 $m = 1$ 成立，则它对于任意的正整数 m 也成立。因此，只要当 $m = 1$ 时 (15.22) 式成立，就可以称随机序列 $\{\xi_n, n = 1, 2, \dots\}$ 具有马氏性，即 $\{\xi_n, n = 1, 2, \dots\}$ 是一个马尔可夫链。

定义 15.8 设 $\{\xi_n, n = 1, 2, \dots\}$ 是一个马氏链。如果等式 (15.22) 右边的条件概率与 n 无关, 即

$$P\{\xi_{n+m} = j \mid \xi_n = i\} = p_{ij}(m) \quad (15.23)$$

则称 $\{\xi_n, n = 1, 2, \dots\}$ 为时齐的马氏链。称 $p_{ij}(m)$ 为系统由状态 i 经过 m 个时间间隔 (或 m 步) 转移到状态 j 的转移概率。(15.23) 式称为时齐性, 它的含义是系统由状态 i 到状态 j 的转移概率只依赖于时间间隔的长短, 与起始的时刻无关。本章介绍的马氏链假定都是时齐的, 因此省略“时齐”二字。

15.4.2 转移概率矩阵及柯尔莫哥洛夫定理

对于一个马尔可夫链 $\{\xi_n, n = 1, 2, \dots\}$ ，称以 m 步转移概率 $p_{ij}(m)$ 为元素的矩阵 $P(m) = (p_{ij}(m))$ 为马尔可夫链的 m 步转移矩阵。当 $m = 1$ 时，记 $P(1) = P$ 称为马尔可夫链的一步转移矩阵，或简称转移矩阵。它们具有下列三个基本性质

(1) 对一切 $i, j \in E$, $0 \leq p_{ij}(m) \leq 1$;

(2) 对一切 $i \in E$, $\sum_{j \in E} p_{ij}(m) = 1$;

(3) 对一切 $i, j \in E$, $p_{ij}(0) = \delta_{ij} = \begin{cases} 1, & \text{当 } i = j \text{ 时,} \\ 0, & \text{当 } i \neq j \text{ 时.} \end{cases}$

当实际问题可以用马尔可夫链来描述时，首先要确定它的状态空间及参数集合，然后确定它的一步转移概率。关于这一概率的确定，可以由问题的内在规律得到，也可以由过去经验给出，还可以根据观测数据来估计。

例 15.8 某计算机机房的一台计算机经常出故障，研究者每隔 15 分钟观察一次计算机的运行状态，收集了 24 小时的数据（共作 97 次观察）。用 1 表示正常状态，用 0 表示不正常状态，所得的数据序列如下

1110010011111110011110111111001111111110001101101
1110110110101111011101111011111110011011111100111

解 设 $X_n (n = 1, \dots, 97)$ 为第 n 个时段的计算机状态，可以认为它是一个时齐马氏链，状态空间 $E = \{0, 1\}$ 。要分别统计各状态一步转移的次数，即 $0 \rightarrow 0$ ， $0 \rightarrow 1$ ， $1 \rightarrow 0$ ， $1 \rightarrow 1$ 的次数，也就是要统计数据字符串中‘00’，‘01’，‘10’，‘11’四个子串的个数。

利用 Matlab 软件，求得 96 次状态转移的情况是

$0 \rightarrow 0$ ，8 次；	$0 \rightarrow 1$ ，18 次；
$1 \rightarrow 0$ ，18 次；	$1 \rightarrow 1$ ，52 次。

因此，一步转移概率可用频率近似地表示为(程序略)

$$p_{00} = P\{X_{n+1} = 0 \mid X_n = 0\} \approx \frac{8}{8+18} = \frac{4}{13},$$

$$p_{01} = P\{X_{n+1} = 1 \mid X_n = 0\} \approx \frac{18}{8+18} = \frac{9}{13},$$

$$p_{10} = P\{X_{n+1} = 0 \mid X_n = 1\} \approx \frac{18}{18+52} = \frac{9}{35},$$

$$p_{11} = P\{X_{n+1} = 1 \mid X_n = 1\} \approx \frac{52}{18+52} = \frac{26}{35}.$$

例 15.9 设一随机系统状态空间 $E = \{1, 2, 3, 4\}$,
记录观测系统所处状态如下

4	3	2	1	4	3	1	1	2	3
2	1	2	3	4	4	3	3	1	1
1	3	3	2	1	2	2	2	4	4
2	3	2	3	1	1	2	4	3	1

若该系统可用马氏模型描述, 估计转移概率 p_{ij} 。

解 记 n_{ij} 是由状态 i 到状态 j 的转移次数，行和

$n_i = \sum_{j=1}^4 n_{ij}$ 是系统从状态 i 转移到其它状态的次数， n_{ij}

和 n_i 的统计数据见表 15.7。一步状态转移概率 p_{ij} 的估

计值 $\hat{p}_{ij} = \frac{n_{ij}}{n_i}$ ，计算得一步状态转移矩阵的估计为（程

序略）

$$\hat{P} = \begin{bmatrix} 2/5 & 2/5 & 1/10 & 1/10 \\ 3/11 & 2/11 & 4/11 & 2/11 \\ 4/11 & 4/11 & 2/11 & 1/11 \\ 0 & 1/7 & 4/7 & 2/7 \end{bmatrix}$$

表 15.7 $i \rightarrow j$ 转移数统计表

	1	2	3	4	行和_{<i>n</i>}_{<i>i</i>}
1	4	4	1	1	10
2	3	2	4	2	11
3	4	4	2	1	11
4	0	1	4	2	7

定理 15.6（柯尔莫哥洛夫—开普曼定理） 设 $\{\xi_n, n = 1, 2, \dots\}$ 是一个马尔可夫链，其状态空间 $E = \{1, 2, \dots\}$ ，则对任意正整数 m, n 有

$$p_{ij}(n+m) = \sum_{k \in E} p_{ik}(n) p_{kj}(m),$$

其中的 $i, j \in E$ 。

定理 15.7 设 P 是一步马氏链转移矩阵（ P 的行向量是概率向量）， $P^{(0)}$ 是初始分布行向量，则第 n 步的概率分布为

$$P^{(n)} = P^{(0)} P^n.$$

例 15.10 若顾客的购买是无记忆的,即已知现在顾客购买情况,未来顾客的购买情况不受过去购买历史的影响,而只与现在购买情况有关。现在市场上供应 A 、 B 、 C 三个不同厂家生产的 50 克袋装味精,用“ $\xi_n = 1$ ”、“ $\xi_n = 2$ ”、“ $\xi_n = 3$ ”分别表示“顾客第 n 次购买 A 、 B 、 C 厂的味精”。显然, $\{\xi_n, n = 1, 2, \dots\}$ 是一个马氏链。若已知第一次顾客购买三个厂味精的概率依次为 0.2, 0.4, 0.4。又知道一般顾客购买的倾向由表 15.8 给出。求顾客第四次购买各家味精的概率。

表 15.8 状态转移概率

		下 次 购 买		
		<i>A</i>	<i>B</i>	<i>C</i>
上次 购买	<i>A</i>	0.8	0.1	0.1
	<i>B</i>	0.5	0.1	0.4
	<i>C</i>	0.5	0.3	0.2

解 第一次购买的概率分布为

$$P^{(1)} = [0.2, 0.4, 0.4],$$

一步状态转移矩阵

$$P = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.5 & 0.1 & 0.4 \\ 0.5 & 0.3 & 0.2 \end{bmatrix},$$

则顾客第四次购买各家味精的概率为

$$P^{(4)} = P^{(1)} P^3 = [0.7004, 0.136, 0.1636].$$

15.4.3 转移概率的渐近性质—极限概率分布

现在考虑，随 n 的增大， P^n 是否会趋于某一固定矩阵？先考虑一个简单例子。

$$\text{转移矩阵 } P = \begin{bmatrix} 0.5 & 0.5 \\ 0.7 & 0.3 \end{bmatrix}, \text{ 当 } n \rightarrow +\infty \text{ 时, } P^n \rightarrow \begin{bmatrix} \frac{7}{12} & \frac{5}{12} \\ \frac{7}{12} & \frac{5}{12} \end{bmatrix}.$$

又若取 $u = \begin{bmatrix} \frac{7}{12} & \frac{5}{12} \end{bmatrix}$ ，则 $uP = u$ ， u^T 为矩阵 P^T 的对应于特征值 $\lambda = 1$ 的特征（概率）向量， u 也称为 P 的不动点向量。哪些转移矩阵具有不动点向量？为此给出正则矩阵的概念。

定义 15.9 一个马氏链的转移矩阵 P 是正则的，
当且仅当存在正整数 k ，使 P^k 的每一元素都是正数。

定理 15.8 若 P 是一个马氏链的正则阵，那么

(1) P 有唯一的不动点向量 W ， W 的每个分量为正。

(2) P 的 n 次幂 P^n (n 为正整数) 随 n 的增加趋于矩阵 \bar{W} ， \bar{W} 的每一行向量均等于不动点向量 W 。

一般地，设时齐马氏链的状态空间为 E ，如果对于所有 $i, j \in E$ ，转移概率 $p_{ij}(n)$ 存在极限

$$\lim_{n \rightarrow \infty} p_{ij}(n) = \pi_j \quad (\text{不依赖于 } i), \text{ 或}$$

$$P(n) = P^n \xrightarrow{(n \rightarrow \infty)} \begin{bmatrix} \pi_1 & \pi_2 & \cdots & \pi_j & \cdots \\ \pi_1 & \pi_2 & \cdots & \pi_j & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \pi_1 & \pi_2 & \cdots & \pi_j & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix},$$

则称此链具有遍历性。又若 $\sum_j \pi_j = 1$ ，则同时称

$\pi = (\pi_1, \pi_2, \cdots)$ 为链的极限分布。

下面就有限链的遍历性给出一个充分条件。

定理 15.9 设时齐马氏链 $\{\xi_n, n = 1, 2, \dots\}$ 的状态空间为 $E = \{a_1, \dots, a_N\}$, $P = (p_{ij})$ 是它的一步转移概率矩阵, 如果存在正整数 m , 使对任意的 $a_i, a_j \in E$, 都有

$$p_{ij}(m) > 0, \quad i, j = 1, 2, \dots, N$$

则此链具有遍历性; 且有极限分布 $\pi = [\pi_1, \dots, \pi_N]$,

它是方程组

$$\pi = \pi P \quad \text{或即} \quad \pi_j = \sum_{i=1}^N \pi_i p_{ij}, \quad j = 1, \dots, N$$

的满足条件

$$\pi_j > 0, \quad \sum_{j=1}^N \pi_j = 1$$

的唯一解。

例 15.11 根据例 15.10 中给出的一般顾客购买三种味精倾向的转移矩阵，预测经过长期的多次购买之后，顾客的购买倾向如何？

解 这个马氏链的转移矩阵满足定理 15.9 的条件，可以求出其极限概率分布。为此，解下列方程组

$$\begin{cases} p_1 = 0.8p_1 + 0.5p_2 + 0.5p_3, \\ p_2 = 0.1p_1 + 0.1p_2 + 0.3p_3, \\ p_3 = 0.1p_1 + 0.4p_2 + 0.2p_3, \\ p_1 + p_2 + p_3 = 1. \end{cases}$$

求得 $p_1 = \frac{5}{7}$, $p_2 = \frac{11}{84}$, $p_3 = \frac{13}{84}$ 。

这说明，无论第一次顾客购买的情况如何，经过长期多次购买以后，A厂产的味精占有市场的 $\frac{5}{7}$ ，B,C两厂产品分别占有市场的 $\frac{11}{84}$ ， $\frac{13}{84}$ 。(程序略)

例 15.12 为适应日益扩大的旅游事业的需要，某城市的甲、乙、丙三个照相馆组成一个联营部，联合经营出租相机的业务。游客可由甲、乙、丙三处任何一处租出相机，用完后，还在三处中任意一处即可。估计其转移概率如表 15.9 所示。

表 15.9 状态转移概率

		还 相 机 处		
		甲	乙	丙
租相 机处	甲	0.2	0.8	0
	乙	0.8	0	0.2
	丙	0.1	0.3	0.6

今欲选择其中之一附设相机维修点，问该点设在哪一个照相馆为最好？

解 由于旅客还相机的情况只与该次租机地点有关,而与相机以前所在的店址无关,所以可用 X_n 表示相机第 n 次被租时所在的店址; “ $X_n = 1$ ”、“ $X_n = 2$ ”、“ $X_n = 3$ ”分别表示相机第 n 次被租用时在甲、乙、丙馆。则 $\{X_n, n = 1, 2, \dots\}$ 是一个马尔可夫链,其转移矩阵 P 由表 15.9 给出。考虑维修点的设置地点问题,实际上要计算这一马尔可夫链的极限概率分布。

状态转移矩阵是正则的，极限概率存在，解方程组

$$\begin{cases} p_1 = 0.2p_1 + 0.8p_2 + 0.1p_3, \\ p_2 = 0.8p_1 + 0.3p_3, \\ p_3 = 0.2p_2 + 0.6p_3, \\ p_1 + p_2 + p_3 = 1. \end{cases}$$

得极限概率 $p_1 = \frac{17}{41}$, $p_2 = \frac{16}{41}$, $p_3 = \frac{8}{41}$ 。

由计算看出，经过长期经营后，该联营部的每架照相机还到甲、乙、丙照相馆的概率分别为 $\frac{17}{41}$ 、 $\frac{16}{41}$ 、 $\frac{8}{41}$ 。由于还到甲馆的照相机较多，因此维修点设在甲馆较好。但由于还到乙馆的相机与还到甲馆的相差不多，若是乙的其它因素更为有利的话，比如，交通较甲方便，便于零配件的运输，电力供应稳定等等，亦可考虑设在乙馆。

15.5 时间序列

15.5.1 平稳性Daniel检验

检验序列平稳性的方法很多，在此介绍其中一种，即 Daniel 检验。Daniel 检验方法建立在 Spearman 相关系数的基础上。

Spearman 相关系数是一种秩相关系数。设 x_1, x_2, \dots, x_n 是从一元总体抽取的容量为 n 的样本，其顺序统计量是 $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ 。若 $x_i = x_{(k)}$ ，则称 k 是 x_i 在样本中的秩，记作 R_i ，对每一个 $i = 1, 2, \dots, n$ ，称 R_i 是第 i 个秩统计量。 R_1, R_2, \dots, R_n 总称为秩统计量。

对于二维总体 (X, Y) 的样本观测数据

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n),$$

可得各分量 X, Y 的一元样本数据 x_1, x_2, \dots, x_n 与
 y_1, y_2, \dots, y_n 。设 x_1, x_2, \dots, x_n 的秩统计量是 R_1, R_2, \dots, R_n ,
 y_1, y_2, \dots, y_n 的秩统计量是 S_1, S_2, \dots, S_n , 当 X, Y 联系比较紧密时, 这两组秩统计量联系也是紧密的。

Spearman 相关系数定义为这两组秩统计量的相关系数，即 **Spearman 相关系数**是

$$q_{xy} = \frac{\sum_{i=1}^n (R_i - \bar{R})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (R_i - \bar{R})^2} \sqrt{\sum_{i=1}^n (S_i - \bar{S})^2}},$$

其中 $\bar{R} = \frac{1}{n} \sum_{i=1}^n R_i$, $\bar{S} = \frac{1}{n} \sum_{i=1}^n S_i$ 。

经过运算，可以证明

$$q_{xy} = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n d_i^2,$$

其中 $d_i = R_i - S_i$, $i = 1, 2, \dots, n$ 。

对于 Spearman 相关系数，作假设检验

$$H_0 : \rho_{XY} = 0, H_1 : \rho_{XY} \neq 0.$$

其中 ρ_{XY} 为总体的相关系数，可以证明，当 (X, Y) 是二元正态总体，且 H_0 成立时，统计量

$$T = \frac{q_{XY} \sqrt{n-2}}{\sqrt{1-q_{XY}^2}}$$

服从自由度为 $n-2$ 的 t 分布 $t(n-2)$ 。

对于给定的显著水平 α ，通过 t 分布表可查到统计量 t 的临界值 $t_{\alpha/2}(n-2)$ ，当 $|t| \leq t_{\alpha/2}(n-2)$ 时，接受 H_0 ；当 $|t| > t_{\alpha/2}(n-2)$ 时，拒绝 H_0 。

对于时间序列的样本 a_1, a_2, \dots, a_n ，记 a_t 的秩为 $R_t = R(a_t)$ ，考虑变量对 (t, R_t) ， $t = 1, 2, \dots, n$ 的 Spearman 相关系数 q_s ，有

$$q_s = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (t - R_t)^2, \quad (15.24)$$

构造统计量

$$T = \frac{q_s \sqrt{n-2}}{\sqrt{1-q_s^2}}.$$

作下列假设检验

H_0 : 序列 X_t 平稳, H_1 : 序列 X_t 非平稳 (存在上升或下降趋势) .

Daniel 检验方法: 对于显著水平 α , 由时间序列 a_t 计算 $(t, R_t), t = 1, 2, \dots, n$ 的 Spearman 秩相关系数 q_s , 若 $|T| > t_{\alpha/2}(n-2)$, 则拒绝 H_0 , 认为序列非平稳。且当 $q_s > 0$ 时, 认为序列有上升趋势; $q_s < 0$ 时, 认为序列有下降趋势。又当 $|T| \leq t_{\alpha/2}(n-2)$ 时, 接受 H_0 , 可以认为 X_t 是平稳序列。

15.5.2 税收收入AR预测模型

1. 问题的提出

税收作为政府财政收入的主要来源，是地区政府实行宏观调控、保证地区经济稳定增长的重要因素。各级政府每年均需预测来年的税收收入以安排财政预算。什么方法能够帮助地方政府有效地预测税收收入？

表 15.10 是某区历年税收数据（单位：亿元）。本节引入现代计量经济学的方法，预测税收收入，为年度税收计划和财政预算提供更有效、更科学的依据。

表 15.10 各年度的税收数据

年份	1	2	3	4	5	6	7
税收	15.2	15.9	18.7	22.4	26.9	28.3	30.5
年份	8	9	10	11	12	13	14
税收	33.8	40.4	50.7	58	66.7	81.2	83.4

2. 模型的构建

从较长的时间来看，经济运行遵循一定的规律，而从短期来看，由于受到宏观政策、市场即期需求变化等不确定因素的影响，预测会有一定的困难。目前，预测经济运行的理论方法有很多，经典的有生长曲线、指数平滑法等等，但这些方法对短期波动的把握性不高。AR 自回归模型在经济预测过程中既考虑了经济现象在时间序列上的依存性，又考虑了随机波动的干扰性，对于经济运行短期趋势的预测准确率较高，是应用比较广泛的一种方法。

作为经济运行的一种重要指标，税收收入具有一定的稳定性和增长性，且与前几年的税收具有一定的关联性，因此可以采用时间序列方法对税收的增长建立预测模型。

记原始是时间序列数据为 a_t ($t = 1, 2, \dots, 14$)，首先检验序列 a_t 是否是平稳的，对显著水平 $\alpha = 0.05$ ，由(15.24)式算得 $q_s = 1$ ，计算得统计量 $T = +\infty$ ，上 $\alpha / 2$ 分位数的值 $t_{\alpha/2}(12) = 2.1788$ ，所以 $|T| > t_{\alpha/2}(n - 2)$ ，故认为序列是非平稳的；因 $q_s > 0$ ，序列有上升趋势。

为了构造平稳序列，对序列 a_t ($t = 1, 2, \dots, 14$) 作一阶差分运算 $b_t = a_{t+1} - a_t$ ，得到序列 b_t ($t = 1, 2, \dots, 13$)。从时间序列 b_t 散点图来看，时间序列是平稳的。可建立如下的自回归模型（AR（2）模型）对 b_t 进行预测，

$$y_t = c_1 y_{t-1} + c_2 y_{t-2} + \varepsilon_t,$$

其中 c_1, c_2 为待定参数， ε_t 为随机扰动项。

3. 模型的求解

根据表 15.10 的数据，采用最小二乘法可计算得出 b_t 的预测模型为

$$y_t = 0.2785y_{t-1} + 0.6932y_{t-2} + \varepsilon_t,$$

利用该模型，求得 $t = 15$ 时，税收的预测值 $\hat{a}_{15} = 94.064$ 。

对于已知数据上述模型的预测相对误差见表 15.11。可以看出该模型的预测精度是较高的。

表 15.11 已知数据的预测值及相对误差

年份	1	2	3	4	5	6	7
税收	15.2	15.9	18.7	22.4	26.9	28.3	30.5
预测 值	15.2	15.9	18.7	19.9651	25.3715	30.7182	31.8093
相对 误差	0	0	0	0.1087	0.0568	0.0854	0.0429
年份	8	9	10	11	12	13	14
税收	33.8	40.4	50.7	58	66.7	81.2	83.4
预测 值	32.0832	36.2442	44.5258	58.1439	67.1731	74.1835	91.2694
相对 误差	0.0508	0.1029	0.1218	0.0025	0.0071	0.0864	0.09436

4. 模型的拓展

由于本案例中第 t 年税收的值与前若干年的值之间具有较高的相关性，所以采用了 AR 模型，在其它情况下，也可采用 MA 模型或者 ARMA 模型等其它时间序列方法。

另外，还可考虑投资、生产、分配结构、税收政策等诸多因素对税收收入的影响，采用多元时间序列分析方法建立关系模型，从而改善税收预测模型，提高预测质量（程序略）。

15.6 插值与拟合

15.6.1 导弹运动轨迹问题

1. 问题的提出

在某次军事演习中，用测距仪对空中的某导弹进行运动轨迹测量。地面上有 3 个测距仪 A_i ($i = 1, 2, 3$)，其中 A_2 位于 A_1 的正北方 4.5km 处， A_3 位于 A_1 与 A_2 的西侧，与 A_1 、 A_2 的距离分别为 $\sqrt{6.25}$ km 与 $\sqrt{13}$ km。测得的数据为每间隔 0.05 秒导弹到 3 个测距仪的距离 (单位:米)，其中最后测得的 10 个数据见表 15.12。试给出 9.5 秒~9.95 秒时间内该导弹的运动轨迹方程。

表 15.12 三个测距仪到导弹的距离数据

时 间 t (秒)	a_1 到 导 弹 距 离 (米)	a_2 到 导 弹 距 离 (米)	a_3 到 导 弹 距 离 (米)
9.5	17675.33388	21839.81626	19851.29886
9.55	18606.75463	22807.17185	20756.90878
9.6	19575.37056	23807.47844	21700.16008
9.65	20580.73101	24840.82393	22680.74614
9.7	21622.49583	25907.33669	23698.43062
9.75	22700.42386	27007.18253	24753.04392
9.8	23814.3622	28140.56181	25844.47924
9.85	24964.23654	29307.70672	26972.68827
9.9	26150.04242	30508.87882	28137.67706
9.95	27371.83745	31744.36674	29339.50183

2. 模型的建立与求解

以 A_1 点作为坐标原点， A_1A_2 所在的射线作为 y 轴的正半轴， A_1, A_2, A_3 所在的平面为 xoy 面，建立三维右手空间坐标系。则 A_1 点的坐标为 $(0, 0, 0)$ ， A_2 点的坐标为 $(0, 4500, 0)$ ，根据距离关系容易求得 A_3 点的坐标为 $(-2000, 1500, 0)$ 。

表 15.12 中的 10 个测量点分别用 $i = 1, 2, \dots, 10$ 编号, 记 d_{ij} ($i = 1, 2, \dots, 10, j = 1, 2, 3$) 为第 i 个测量点到第 j 个测距仪的距离, 设第 i 个测量点的坐标为 (x_i, y_i, z_i) , 由点 (x_i, y_i, z_i) 到三个测距仪的距离关系建立如下非线性方程组

$$\begin{cases} x_i^2 + y_i^2 + z_i^2 = d_{i1}^2, \\ y_i^2 + (y_i - 4500)^2 + z_i^2 = d_{i2}^2, \\ (x_i + 2000)^2 + (y_i - 1500)^2 + z_i^2 = d_{i3}^2. \end{cases} \quad (i = 1, 2, \dots, 10)$$

(15.25)

求解 (15.25) 式的方程组即可求得 10 个观察点的三维坐标。

下面建立导弹运动轨迹的参数方程

$$\begin{cases} x = x(t), \\ y = y(t), \\ z = z(t), \end{cases}$$

由于只有 10 个观测点，必须进行插值，这里我们使用三次样条函数进行插值，利用 Matlab 程序就可以求得导弹运动的轨迹，把插值的轨迹方程全部写出来太繁杂，这里只给出最后一个区间，

即最后两个观测点之间的轨迹参数方程为（程序略）

$$\begin{cases} x = 0.0129(t - 9.90)^3 - 30.1605(t - 9.90)^2 \\ \quad - 775.4497(t - 9.90) + 6520.4573, \\ y = -704.44(t - 9.90)^3 - 6265.95(t - 9.90)^2 \\ \quad - 25285.76(t - 9.90) - 25190.77, \\ z = -0.3452(t - 9.90)^3 - 4.9089(t - 9.90)^2 \\ \quad - 898.0054(t - 9.90) + 2594.8503. \end{cases}$$

其中 $t \in [9.9, 9.95]$ 。

15.6.2 录像机计数器的用途

1. 问题的提出

老式的录像机上都有计数器，而没有计时器。经试验，一盘标明 180 分钟的录像带从头走到尾，时间用了 184 分，计数器读数从 0000 变到 6061，另外还有一批测试数据，所有的测试数据见表 15.13。

在一次使用中录像带已经转过大半，计数器读数为 4450，问剩下的一段还能否录下 1 小时的节目？

表 15.13 时间 t 和计数器 n 的测量数据

t (分钟)	0	20	40	60	80
n	0	1141	2019	2760	3413
t (分钟)	100	120	140	160	184
n	4004	4545	5051	5525	6061

2. 问题分析

计数器的读数是怎样变化的，它的增长为什么先快后慢，回答这个问题需要了解计数器的简单工作原理（图 15.1）。

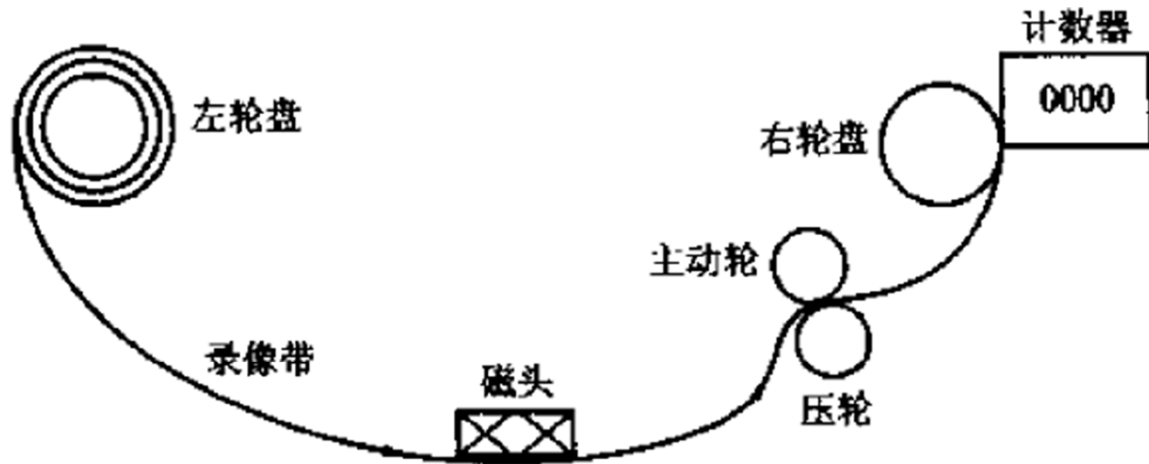


图 15.1 录像机计数器工作原理示意图

录像带有两个轮盘，一开始录像带缠满的那个轮盘不妨称为左轮盘，另一个为右轮盘。计数器与右轮盘的轴相连，其读数与右轮转动的圈数成正比。开始时右轮盘是空的，读数为 0000，随着带子从左向右运动，右轮盘半径增加，使得转动越来越慢，计数器读数的增长也就越来越慢。

在录像带的转动过程中，与微型电机相连的主动轮的转速当然是不变的，录像带靠压轮压在主动轮上，所以录像带的运动速度（线速度）为常数，而右轮盘随着半径的增加，其转速当然越来越慢了。

我们要找出计数器读数 n 与录像带转过时间 t 之间的关系，即建立一个数学模型 $t = f(n)$ 。

3. 模型假设

- (1) 录像带的线速度是常数 v ;
- (2) 计数器读数 n 与右轮盘转的圈数 m 成正比,
 $m = kn$, k 为比例系数;
- (3) 录像带的厚度(加上缠绕时两圈间的空隙)是常数 w , 空右轮盘半径为 r ;
- (4) 初始时刻 $t = 0$ 时 $n = 0$ 。

4. 模型建立

建立 t 与 n 之间的关系有多种途径。

方法一

计算缠绕在右轮盘上的录像带的长度。当右轮盘转到第 i 圈时，其半径为 $r + wi$ ，周长为 $2\pi(r + wi)$ ， m 圈的总长度恰等于录像带转过的长度 vt ，即

$\sum_{i=1}^m 2\pi(r + wi) = vt$ ，代入 $m = kn$ ，容易算出

$$t = \frac{\pi wk^2}{v} n^2 + \frac{2\pi rk + \pi wk}{v} n, \quad (15.26)$$

这就是我们需要的数学模型。

方法二

考察右轮盘面积的增加,它应该等于录像带转过的长度与厚度的乘积,即

$$\pi[(r + wkn)^2 - r^2] = wvt,$$

可以算出

$$t = \frac{\pi wk^2}{v} n^2 + \frac{2\pi rk}{v} n. \quad (15.27)$$

模型 (15.26) 与 (15.27) 有微小的差别。考虑到 w 比 r 小得多, 使用模型 (15.27) 即可。

方法三

用微元分析法，考察 t 到 $t + dt$ 时间内录像带在右轮盘缠绕圈数从 m 变化到 $m + dm$ ，有

$$2\pi(r + mw)dm = vdt,$$

即

$$2\pi(r + knw)kdn = vdt,$$

再加上初始条件，建立微分方程的初值问题

$$\begin{cases} \frac{dt}{dn} = \frac{2\pi rk}{v} + \frac{2\pi wk^2}{v}n, \\ t|_{n=0} = 0. \end{cases}$$

同样可以得到式 (15.27)。

实际上，从建模的目的看，如果把（15.27）式改记作

$$t = an^2 + bn, \quad (15.28)$$

那么只需要确定 a, b 两个参数即可进行 n 和 t 之间的计算。

5. 模型的求解

使用表 15.13 中的数据，利用最小二乘法估计参数 a, b ，求得 $a = 2.61 \times 10^{-6}$, $b = 1.45 \times 10^{-2}$ ，代入 (15.28) 式，算得 $n = 4450$ 时， $t = 116.4$ 分，剩下的录像带能录 $184 - 116.4 = 67.6$ 分钟的节目（程序略）。

15.7 神经元网络

人工神经网络是国际学术界十分活跃的前沿研究领域，在控制与优化、预测与管理、模式识别与图像处理、通信等方面得到了十分广泛的应用。

下面简单介绍 BP (Back Propagation) 神经网络和径向基函数 (Radial Basis Function, 简称 RBF) 神经网络的原理，及其在预测中的应用。

15.7.1 BP神经网络

1 BP神经网络拓扑结构

BP 神经网络是一种具有三层或三层以上的多层神经网络，每一层都由若干个神经元组成，如图 15.2 所示，它的左、右各层之间各个神经元实现全连接，即左层的每一个神经元与右层的每个神经元都有连接，而上下各神经元之间无连接。BP 神经网络按有导师学习方式进行训练，当一对学习模式提供给网络后，其神经元的激活值将从输入层经各隐含层向输出层传播，在输出层的各神经元输出对应于输入模式的网络响应。

然后，按减少希望输出与实际输出误差的原则，从输出层经各隐含层、最后回到输入层逐层修正各连接权。由于这种修正过程是从输出到输入逐层进行的，所以称它为“误差逆传播算法”。随着这种误差逆传播训练的不断修正，网络对输入模式响应的正确率也将不断提高。

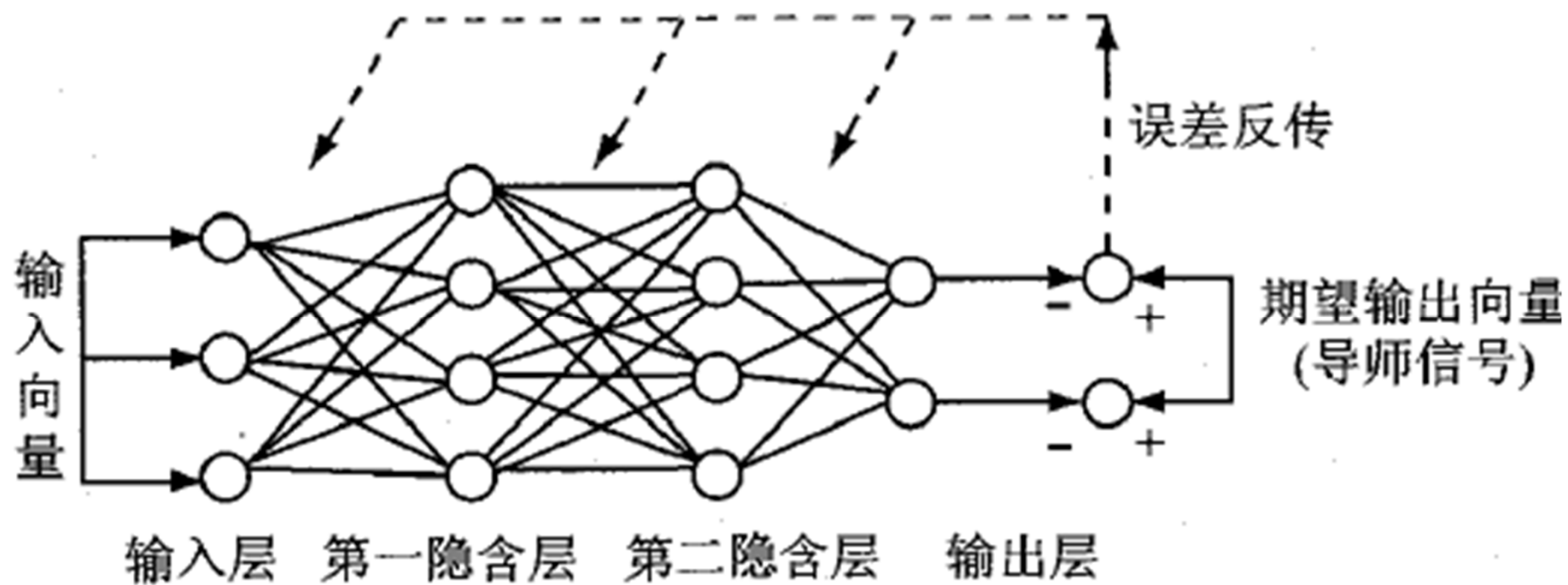


图 15.2 BP 网络模型结构

2 BP神经网络训练

为了使 BP 神经网络具有某种功能，完成某项任务，必须调整层间连接权值和节点阈值，使所有样品的实际输出和期望输出之间的误差稳定在一个较小的值以内。

一般地，BP 网络的学习算法描述为如下步骤

(1) 初始化网络及学习参数，如设置网络初始权矩阵，学习因子等。

(2) 提供训练模式，训练网络，直到满足学习要求。

(3) 前向传播过程：对给定训练模式输入，计算网络的输出模式，并与期望模式比较，若有误差，若执行步骤(4)，否则，返回步骤(2)。

(4) 反向传播过程：计算同一层单元的误差，修正权值和阈值，返回步骤 (2)。

网络的学习是通过用给定的训练集训练而实现的。通常用网络的均方差误差来定量地反映学习的性能。一般地，当网络的均方差误差低于给定值时，则表明对给定训练集学习已满足要求了。

15.7.2 RBF神经网络

1 RBF网络结构

RBF 神经网络有很强的逼近能力、分类能力和学习速度。其工作原理是把网络看成对未知函数的逼近，任何函数都可以表示成一组基函数的加权和，也即选择各隐层神经元的传输函数，使之构成一组基函数来逼近未知函数。RBF 人工神经网络由一个输入层、一个隐含层和一个输出层组成。RBF 神经网络的隐层基函数有多种形式，常用函数为高斯函数，设输入层的输入为 $X = [x_1, x_2, \cdots, x_n]$ ，实际输出为

$$Y = [y_1, y_2, \cdots, y_p]。$$

输入层实现从 $X \rightarrow R_i(X)$ 的非线性映射，输出层实现从 $R_i(X) \rightarrow y_k$ 的线性映射，输出层第 k 个神经元网络输出为

$$\hat{y}_k = \sum_{i=1}^m w_{ik} R_i(X), \quad k = 1, \dots, p, \quad (15.29)$$

式中， n 为输入层节点数； m 为隐含层节点数； p 为输出层节点数； w_{ik} 为隐含层第 i 个神经元与输出层第 k 个神经元的连接权值；

$R_i(X)$ 为隐含层第 i 个神经元的作用函数，如下式所示

$$R_i(X) = \exp(-\|X - C_i\|^2 / 2\sigma_i^2), \quad i = 1, \dots, m, \quad (15.30)$$

式中， X 为 n 维输入向量； C_i 为第 i 个基函数的中心，与 X 具有相同维数的向量， σ_i 为第 i 个基函数的宽度， m 为感知单元的个数（隐含层节点数）。 $\|X - C_i\|$ 为向量 $X - C_i$ 的范数，它通常表示 X 与 C_i 之间的距离；

$R_i(X)$ 在 C_i 处有一个唯一的最大值，随着 $\|X - C_i\|$ 的增大， $R_i(X)$ 迅速衰减到零。对于给定的输入，只有一小部分靠近 X 的中心被激活。当确定了 RBF 网络的聚类中心 C_i 、权值 w_{ik} 及 σ_i 以后，就可求出给定某一输入时，网络对应的输出值。

2 RBF网络学习算法

在 RBF 网络中，隐层执行的是一种固定不变的非线性变换， C_i, σ_i, w_{ik} 需通过学习和训练来确定，一般分为 3 步进行。

(1) 确定基函数的中心 C_i 。利用一组输入来计算 m 个 C_i ， $i = 1, 2, \dots, m$ ，使 C_i 尽可能均匀地对数据抽样，在数据点密集处 C_i 也密集。一般采用“ K 均值聚类法”。

(2) 确定基函数的宽度 σ_i 。基函数中心 C_i 训练完成后，可以求得归一化参数，即基函数的宽度 σ_i ，表示与每个中心相联系的子样本集中样本散布的一个测度。常用的是令其等于基函数中心与子样本集中样本模式之间的平均距离。

(3) 确定从隐含层到输出层的连接权值 w_{ik} ,
RBF 连接权 w_{ik} 的修正可以采用最小均方差误差测
度准则进行。

15.7.3 基于神经网络的年径流预报实例

1. 网络学习样本的建立

现有某水库实测径流资料和相应的前期 4 个预报因子实测数据见表 15.14，其中 4 个预报因子分别为水库上一年 11~12 月的总降雨量 x_1 （单位 mm），当年 1,2,3 月的总降雨量 x_2, x_3, x_4 。在本例应用中将这 4 个预报因子作为输入，年径流量 y （单位 m^3/s ）为输出，构成 4 个输入 1 个输出的网络，将前 19 个实测数据作为训练样本集，后 1 个实测数据作为预测检验样本。

表 15.14 某水库实测年径流量与因子特征值

序号	x_1	x_2	x_2	x_2	y
1	15.6	5.6	3.5	25.5	22.9
2	27.8	4.3	1.0	7.7	23.4
3	35.2	3.0	38.1	3.7	36.8
4	10.2	3.4	3.5	7.4	22.0
5	29.1	33.2	1.6	24.0	6.4
6	10.2	11.6	2.2	26.7	29.4
7	35.4	4.1	1.3	7.0	26.2
8	8.7	3.5	7.5	5.0	20.9
9	25.4	0.7	22.2	35.4	26.5
10	15.3	6.0	2.0	17.5	37.3
11	25.9	1.2	9.0	3.3	22.8
12	64.3	3.7	4.6	4.8	19.8

2. 原始数据的预处理

采用 (15.31) 式分别对样本的输入、输出数据进行规格化处理

$$\tilde{t} = \frac{2(t - t_{\min})}{t_{\max} - t_{\min}} - 1, \quad (15.31)$$

式中 t 为规格化前的变量, t_{\max} 和 t_{\min} 分别为 t 的最大值和最小值, \tilde{t} 为规格化后的变量。

Matlab 中提供了对数据进行规格化处理的函数

[tn, ps]=mapminmax(t)

相应的逆处理函数

t=mapminmax('reverse', tn, ps)

执行的算法是

$$t = 0.5(\tilde{t} + 1) \cdot (t_{\max} - t_{\min}) + t_{\min}.$$

3. 网络的训练

利用的 Matlab 提供神经网络工具箱实现人工神经网络的功能十分方便。由于年径流预报中自变量有 4 个，因变量有 1 个，输入神经元的个数取为 4，输出神经元的个数取为 1，中间隐含层神经元的个数，BP 网络需要根据经验取定，RBF 网络会在训练过程中自适应地取定。

BP网络存在一些缺点，如收敛速度慢，网络易陷于局部极小，学习过程常常发生振荡。对于本案例的预测，BP网络隐含层神经元个数取为4时，计算结果相对稳定，隐含层神经元个数取为其它值时，运行结果特别不稳定，每一次的运行结果相差很大。

利用Matlab工具箱，求得对于第20个样本点，RBF网络的预测值为26.7693，相对误差为18.39%，BP网络的运行结果每次都有很大的不同，通过计算结果可以看出，RBF网络模型的预测结果要好于BP网络模型的预测结果。（程序略）