# 第15章 预测方法

数学建模算法与应用

# 目录


- 15.1 微分方程模型
- 15.2 灰色预测模型
- 15.3 回归分析预测方法
- 15.4 差分方程
- 15.5 马尔可夫预测
- 15.6 时间序列
- 15.7 插值与拟合
- 15.8 神经元网络

# 15.1 微分方程模型

例15.1 美日硫磺岛战斗模型

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Lanchester正规战斗模型进行了验证，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。

# 模型建立

- 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：

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

- 美军战地记录给出增援u(t)为：

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

# 确定参数

- 离散化：

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

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

- 代人$\sum_{\tau=1}^{36}A(\tau)=2037000$，可确定参数：

$$b=\frac{21500-0}{2037000}=0.0106$$

$$a=\frac{\sum_{\tau=1}^{t}u(\tau)-A(36)}{\sum_{\tau=1}^{t}J(\tau)}=0.0544$$

# 15.2 灰色预测模型

- 灰色系统：

白色系统是指系统内部特征是完全已知的；

黑色系统是指系统内部信息完全未知的；

而灰色系统是介于白色系统和黑色系统之间的一种系统，灰色系统其内部一部分信息已知，另一部分信息未知或不确定。

# 15.2 灰色预测模型

- 灰色预测：

　　灰色预测，是指对系统行为特征值的发展变化进行的预测，对既含有已知信息又含有不确定信息的系统进行的预测，也就是对在一定范围内变化的、与时间序列有关的灰过程进行预测。尽管灰过程中所显示的现象是随机的、杂乱无章的，但毕竟是有序的、有界的，因此得到的数据集合具备潜在的规律。灰色预测是利用这种规律建立灰色模型对灰色系统进行预测。

# 15.2 灰色预测模型

- GM(1,1)模型：

目前使用最广泛的灰色预测模型就是关于数列预测的一个变量、一阶微分的GM(1,1)模型。它是基于随机的原始时间序列，经按时间累加后所形成的新的时间序列呈现的规律可用一阶线性微分方程的解来逼近。

经一阶线性微分方程的解逼近所揭示的原始时间序列呈指数变化规律。因此，当原始时间序列隐含着指数变化规律时，灰色模型GM(1,1)的预测是非常成功的。

# 基本模型：GM(1,1)

- 2.1 GM(1,1)的一般形式；
- 2.2 辩识算法；
- 2.3 预测值的还原；
- 2.4 GM(1,1)模型的检验。

# GM(1,1)的一般形式

设某一预测对象的非负单调原始数据列为：

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

为了建立灰色预测模型：

1、对 $x^{(0)}$ 进行一次累加，生成一次累加序列 $x^{(1)}$ ：

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

这里：

$$x^{(1)}(t)=x^{(1)}(1)+\cdots+x^{(1)}(t);$$

2、建立GM(1,1)：

即对 $x^{(1)}$ 建立微分方程：$\dfrac{dx^{(1)}}{dt}+ax^{(1)}=u$

# 辩识算法

为了确定模型中的发展系数$a$与灰色作用量$u$，记：

$$a=\begin{bmatrix} a \\ u \end{bmatrix}$$

3、取：

$$B=\begin{bmatrix} -(x^{(1)}(1)+x^{(1)}(2))/2 & 1 \\ -(x^{(1)}(2)+x^{(1)}(3))/2 & 1 \\ \vdots & \vdots \\ -(x^{(1)}(n)+x^{(1)}(n-1))/2 & 1 \end{bmatrix} \qquad Y=\begin{bmatrix} x^{(1)}(2) \\ x^{(1)}(3) \\ \cdots \\ x^{(1)}(n) \end{bmatrix}$$

4、于是：

$$a=\begin{bmatrix} a \\ u \end{bmatrix}=(B^T B)^{-1} B^T Y$$

# 预测值的还原

将参数代人GM模型，可以得到的一次累加量$x^{(1)}$的预测值，再将GM模型所得数据经过累减，还原为$x^{(0)}$的预测值。

5、对微分方程求解：

$$x^{(1)}(t)=(x^{(0)}(1)-\frac{u}{a})e^{-a(t-1)}+\frac{u}{a}$$

6、还原为：

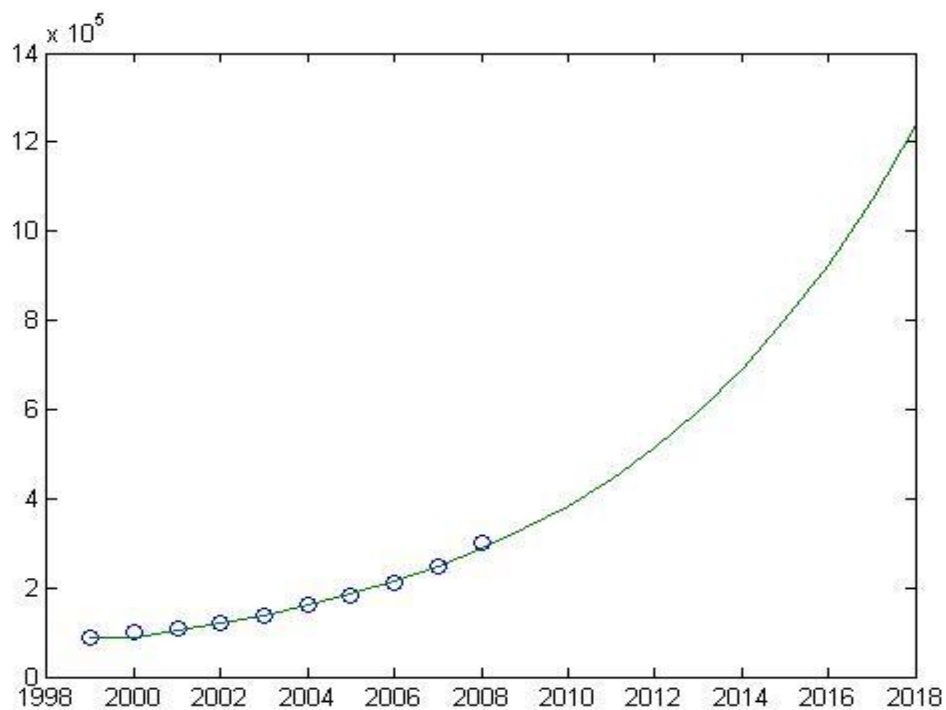
$$x^{(0)}(t)=x^{(1)}(t)-x^{(1)}(t-1)$$

# GM(1,1)模型的检验

- 相对误差
- 关联度
- 均方差比值
- 小误差概率

# 应用举例

设有某公司1999 – 2008年的利润为（单位：万元/年）：{89677,99215,109655,120333,135823,159878,182321,209407,246619,300670}，现在预测该公司未来几年的利润情况。



# 附录

```
clear;clc
syms a b;
c=[a b];

A=[89677,99215,109655,120333,135823,159878,182321,209407,246619,300670];

B=cumsum(A); n=length(A);
for i=1:(n-1)
   C(i)=(B(i)+B(i+1))/2;
end
D=A;D(1)=[];
D=D';
E=[-C;ones(1,n-1)];
c=inv(E*E')*E*D; c=c';
a=c(1);b=c(2);

F=[];F(1)=A(1);
for i=2:(n+10)
   F(i)= (A(1)-b/a)/exp(a*(i-1)) + b/a;
end
G=[];G(1)=A(1);
for i=2:(n+10)
   G(i)=F(i)-F(i-1);
end

t1=1999:2008;
t2=1999:2018
G
plot(t1,A,'o',t2,G)
```

# 特点

较之其它常规的时间序列预测法,应用灰色模型进行预测有以下显著的特点。

(1)灰色模型是一种长期预测模型，将预测系统中的随机元素作为灰色数据进行处理，而找出数据的内在规律。进行预测所需原始数据量小，预测精度较高，无须像其它预测法要么需要数据量大且规律性强，要么需要凭经验给出系数。

(2)理论性强，计算方便。

(3)用有限的表征系统行为特征的外部元素，分析系统的内在规律。

(4) 适用性强。用灰色模型既可对周期性变化的系统行为进行预测，亦可对非周期性变化的系统行为进行预测；既可进行宏观长期的预测，亦可用于微观短期的预测。

# 推广

- DGM（2,1）
- GM（2,1）
- 灰色Verhulst模型

# 15.3 回归分析预测方法

- 通常曲线拟合问题：根据得到的若干有关变量的一组数据，寻找因变量与（一个或几个）自变量之间的一个函数，使这个函数对那组数据拟合得最好。函数的形式可以由经验、先验知识或对数据的直观观察决定，要作的工作是由数据用最小二乘法计算函数中的待定系数。
- 从数理统计的观点看，所涉及的都是随机变量，我们根据一个样本计算出的那些系数，只是它们的一个（点）估计，应该对它们作区间估计或假设检验；也可以用方差分析方法对模型的误差进行分析，对拟合的优劣给出评价。
- 简单地说，回归分析就是对拟合问题作的统计分析。

# 回归分析研究的几个问题

- （i）建立因变量 $y$ 与自变量 $x_1, x_2, \ldots, x_n$ 之间的回归模型；
- （ii）对回归模型的可信度进行检验；
- （iii）判断每个自变量 $x_i$ 对 $y$ 的影响是否显著；
- （iv）诊断回归模型是否适合这组数据；
- （v）利用回归模型对 进行预报或控制。

# 多元线性回归

- 1.1 模型
- 1.2 参数估计
- 1.3 统计分析
- 1.4 回归模型的假设检验
- 1.5 回归系数的假设检验和区间估计
- 1.6 利用回归模型进行预测
- 1.7 Matlab实现

# 1.1 模型

回归分析中，自变量$x=(x_1,x_2,\cdots,x_m)$ 是影响因变量$y$ 的主要因素，是人们能控制或能观察的，但$y$而还受到随机因素的影响，可以合理地假设这种影响服从零均值的正态分布。于是：

$$\begin{cases} y=\beta_0+\beta_1 x_1+\cdots+\beta_m x_m+\varepsilon \\ \varepsilon \sim N(0,\sigma^2) \end{cases}$$ 其中$\sigma$未知。

- 现得到$n$个独立观测数据$(y_i, x_{i1},\cdots,x_{im})$，$i=1,\cdots,n, n>m$，由上式可得：

$$\begin{cases} y_i=\beta_0+\beta_1 x_{i1}+\cdots+\beta_m x_{im}+\varepsilon_i \\ \varepsilon_i \sim N(0,\sigma^2), \quad i=1,\cdots,n \end{cases}$$

# 记号

- 记

$$X=\begin{bmatrix}1 & x_{11} & \cdots & x_{1m}\\ \vdots & \vdots & \cdots & \vdots\\ 1 & x_{n1} & \cdots & x_{nm}\end{bmatrix} \qquad Y=\begin{bmatrix}y_1\\ \vdots\\ y_n\end{bmatrix}$$

$$\varepsilon=[\varepsilon_1 \quad \cdots \quad \varepsilon_n]^T \qquad \beta=[\beta_0 \quad \beta_1 \quad \cdots \quad \beta_m]^T$$

- 可简化为：

$$\begin{cases}Y=X\beta+\varepsilon\\ \varepsilon\sim N(0,\sigma^2)\end{cases}$$

# 1.2 参数估计

利用最小二乘法估计模型中的参数。

- 误差平方和：$Q(\beta)=\sum_{i=1}^{n}\varepsilon_i^2=(Y-X\beta)^T(Y-X\beta)$， 求$\beta$使$Q(\beta)$最小，所得$\beta$最小二乘估计，记作$\hat{\beta}$，可以推出：

$$\hat{\beta}=(X^TX)^{-1}X^TY$$

将$\hat{\beta}$代回原模型，得到$y$的估计值：

$$\hat{y}=\hat{\beta}_0+\hat{\beta}_1x_1+\cdots+\hat{\beta}_mx_m$$

- 拟合误差$e=Y-\hat{Y}$称为残差，可作为随机误差$\varepsilon$的估计。
- 残差平方和（或剩余平方和），即：

$$Q=Q(\hat{\beta})=\sum_{i=1}^{n}e_i^2=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$$

# 1.3 统计分析

理论分析中给出了下列结果：

- 1、$\hat{\beta}$是$\beta$的线性无偏最小方差估计；
- 2、$\hat{\beta}$服从正态分布：

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1})$$

- 3、对残差平方和$Q$，$EQ = (n-m-1)\sigma^2$，且 $\frac{Q}{\sigma^2} \sim \chi^2(n-m-1)$

由此得到的无偏估计 $s^2 = \frac{Q}{n-m-1} = \hat{\sigma}^2$

$s^2$是剩余方差（残差的方差）。

- 4、对总平方和 $S = \sum_{i=1}^{n} (y_i - \overline{y})^2$ 进行分解，有

$$S = Q + U = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2$$

其中 $Q$是由定义的残差平方和，反映随机误差对 $y$的影响，$U$ 称为回归平方和，反映自变量对 $y$ 的影响。

# 1.4 回归模型的假设检验

因变量$y$与自变量 $x_1,\cdots,x_m$之间是否存在如模型所示的线性关系是需要检验的。

显然，若所有的$|\hat{\beta}_j|\,(j=1,\cdots,m)$都很小，线性关系就不明显。

- 令原假设为：

$$H_0:\beta_j=0(j=1,\cdots,m)$$

当$H_0$成立时，由分解式定义的 $U,Q$满足：

$$F=\frac{U/m}{Q/(n-m-1)}\sim F(m,n-m-1)$$

在显著性水平 $\alpha$下有上$\alpha$分位数 $F_{1-\alpha}(m,n-m-1)$，

若 $F<F_{1-\alpha}(m,n-m-1)$，接受 $H_0$；否则，拒绝。

# 1.5 回归系数的假设检验和区间估计

当上面的$H_0$被拒绝时，$\beta_j$不全为零，但是不排除其中若干个等于零。所以应进一步作如下 $m+1$ 个检验：

$$H_0^{(j)}:\beta_j=0$$

- 当 $H_0^{(j)}$成立时

$$t_j=\frac{\hat{\beta}_j/\sqrt{c_{jj}}}{\sqrt{Q/(n-m-1)}}\sim t(n-m-1)$$

对给定的 $\alpha$，若$|t_j|<t_{1-\frac{\alpha}{2}}(n-m-1)$，接受 $H_0^{(j)}$ ；否则，拒绝。

# 1.6 利用回归模型进行预测

- 预测公式：

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \cdots + \hat{\beta}_m x_{0m}$$

- 剔除异常数据

# 推广：

- 多元二项式回归

- 非线性回归

# 15.4 差分方程

# 15.5 马尔可夫预测

- 15.5.1 马尔可夫链的定义
- 15.5.2 转移概率矩阵及柯尔莫哥洛夫定理
- 15.5.3 转移概率的渐近性质：极限概率分布

# 15.5.1 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。

- 定义【马氏链】设$\{\xi_n, n=1,2,\cdots\}$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 $i, j, i_k \in E(k=1,\cdots,n-1)$ ，有

$$P\{\xi_{n+m}=j \mid \xi_n=i, \xi_{n-1}=i_{n-1},\cdots,\xi_1=i_1\}=P\{\xi_{n+m}=j \mid \xi_n=i\}$$

则称$\{\xi_n, n=1,2,\cdots\}$为一个马尔可夫链（简称马氏链）。

- 事实上，只要 $m=1$ 时上式成立，就可以称随机序列 $\{\xi_n, n=1,2,\cdots\}$ 具有马氏性，即它是一个马尔可夫链。

# 时齐的马氏链

- 定义【时齐的马氏链】设 $\{\xi_n, n=1,2,\cdots\}$是一个马氏链。如果马氏链等式右边的条件概率与无关，即：

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

则称$\{\xi_n, n=1,2,\cdots\}$为时齐的马氏链。并称$p_{ij}(m)$为系统由状态$i$ 经过$m$个时间间隔（或 步）转移到状态$j$ 的转移概率。

- 上式称为时齐性。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。
- 通常马氏链假定是时齐的。

# 15.5.2 转移概率矩阵及柯尔莫哥洛夫定理

- 对于一个马尔可夫链， 称矩阵$P(m)=(p_{ij}(m))$为马尔可夫链的$m$步转移矩阵。
- 特别地， 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。

 转移矩阵具有下列三个基本性质：

- （i）$0\le p_{ij}(m)\le 1$ ；
- （ii） $\sum_{j\in E} p_{ij}(m)=1$；
- （iii） $p_{ij}(0)=\delta_{ij}=\begin{cases}1, & \text{当} i=j \text{时} \\ 0, & \text{当} i\neq j \text{时}\end{cases}$ 。

# 转移概率的确定

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，接着就要确定它的一步转移概率。

转移概率的确定，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、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：

  1110010011111110011101111110011111111100011011011101
  1011010111101110111101111110011011111100111

# 柯尔莫哥洛夫—开普曼定理

- 定理1（柯尔莫哥洛夫—开普曼定理）设$\{\xi_n, n=1,2,\cdots\}$是一个马尔可夫链，其状态空间 $E=\{1,2,\cdots\}$，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 $i,j\in E$ 。

# 概率分布

- 定理：设$P$是一个马氏链转移矩阵（ $P$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、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 三个不同厂家$A$、$B$、$C$ 生产的50克袋状味精，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”。这是一个马氏链。

  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为：

$$\begin{matrix} 0.8 & 0.1 & 0.1 \\ 0.5 & 0.1 & 0.4 \\ 0.5 & 0.3 & 0.2 \end{matrix}$$

求顾客第四次购买各家味精的概率。

-

# 15.5.3 转移概率的渐近性质

- 定义【正则】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：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$W$，它的每个分量为正。

  （2）$P$ 的$n$次幂（$n$ 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 $w$ 。

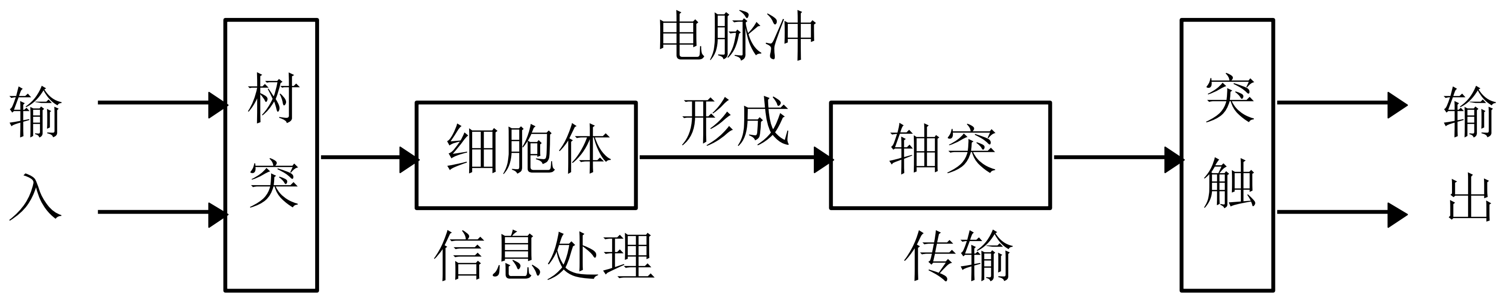
# 预测市场占有率

- 例 根据上例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？

# 15.6 时间序列

# 15.7 插值与拟合

# 15.8 神经元网络



# 评价

- 一般地，ANN与经典计算方法相比并非优越。只有当常规方法解决不了或效果不佳时ANN方法才显示出优越性。尤其对问题的机理不甚了解或者用传统数学模型方法表示的系统，如：特征提取和预测等问题，ANN往往是最有利的工具。另一方面, ANN对处理大量原始数据而不能用规则或公式描述的问题，表现出极大的灵活性和自适应性。
- 人工神经网络以其具有自学习、自组织、较好的容错性和优良的非线性逼近能力，受到众多领域学者的关注。在实际应用中，80%～90%的人工神经网络模型是采用误差反传算法或其变化形式的网络模型（简称BP网络），目前主要应用于函数逼近、模式识别、分类和数据压缩或数据挖掘。