

# 组合样本指标的模糊多级判别与预测

陈治典

(应用数学系)

## 摘 要

本文是将模糊数学方法与多元统计方法相结合的一种尝试,所述方法能更充分地利用样本指标(线性的及非线性的)信息,具有处理模糊信息的能力,不受样本数据概率分布的限制,因而提高了多级判别与预报效果,可作趋势及定值预报,並更广泛地应用于各种多级判别与预测问题。

## 一、引 言

常用的多元统计方法(如回归分析、Fisher 判别、Bayes 意义下的逐步判别等方法)往往要求样本数据满足正态线性条件和各类总体的协方差阵相等,而许多实际问题常不满足(或难以判定是否满足)这些条件。如何突破这些条件的限制,从而扩大多元分析的应用范围,这是近年来数理统计工作者致力研究的问题之一。另一方面,客观世界充满着模糊性,对许多样本往往不能清晰地分类,所谓“因子的预报能力”、“合理分类”、“效果好坏”等都是模糊概念,处理这些模糊概念用模糊数学方法更为合理和有效。本文力图将模糊数学方法与多元统计方法结合起来,以期能更合理、更有效地作出多级判别与预测,且使所建立的方法能适用于任意分布的样本数据。为了更充分地利用大量线性的及非线性的因子(或样本指标)的信息,我们首先对待选因子及预报对象的数据进行秩变换<sup>[1]</sup>,用适当的方法(聚类分析或试探组合)将每5个因子(或样本指标)组成一组,取其秩数之和作为一个新因子(样本指标),从而显著地提高了新因子与预报对象的相关性。然后用统计的办法求出各“秩和”因子对预报对象的预报能力,同时根据各因子当前所属级别求出单因子评判矩阵,再用模糊综合评判的方法<sup>[2]</sup>(本文已作改进,采用滑动权系数)评定预报对象的级别(即趋势),最后用模糊模式识别的方法作出定值预报。与文[3]相比,本文所述方法效果更好。

## 二、方法步骤

### 1. 确定预报对象及挑选预报因子

根据实际问题确定预报对象,记作 $y$ 。按专业知识及实践经验寻找一批与 $y$ 有关的因子

本文1987年5月23日收到。

(样本指标), 用秩相关系数检验法<sup>[1]</sup>从中选出与  $y$  关系较为密切的因子  $x_1, x_2, \dots, x_l$ . 设  $y$  与各因子都有  $n$  个历史观测值

$$y_1, y_2, \dots, y_n; \quad x_{11}, x_{12}, \dots, x_{1n}; \quad i = 1, 2, \dots, l$$

设各因子的当前值为  $x_{10}, x_{20}, \dots, x_{l0}$ .

## 2. 构造“秩和”新因子

将  $y$  的历史数据按从小到大的顺序排列定秩, 以  $R(y_j)$  表示  $y_j$  之秩 ( $j = 1, 2, \dots, n$ ). 若因子与  $y$  正相关, 则将其历史数据按从小到大的顺序排列定秩; 若因子与  $y$  负相关, 则将其历史数据按从大到小的顺序排列定秩. 至于因子当前值则以它与该因子历史数据中最接近的那个数据之秩作为它的秩的估计值, 用聚类分析或试探组合的办法将每 5 个因子组成一组, 取其秩数之和作为一个新因子, 共可得  $m = [l/5]$  个新因子, 记作  $x'_{11}, x'_{12}, \dots, x'_{1m}$ . 新因子的历史观测值及当前值分别记作  $x'_{ij}$  及  $x'_{i0}$  ( $i = 1, 2, \dots, m; j = 1, 2, \dots, n$ ). 每个新因子都集中了 5 个因子的信息, 并且由于经过了秩变换, 原来因子跟  $y$  的非线性关系也线性化了, 故新因子跟  $y$  之秩的相关系数都比原因子跟  $y$  的相关系数有显著提高 (参看后面实例).

## 3. 分级并简化样本数据

根据  $y$  的历史数据的差异, 将  $y$  的数据适当地分成  $s$  级, 使各级之间有明显的差异, 记作  $Y^{(1)}, Y^{(2)}, \dots, Y^{(s)}$ :

$$Y^{(k)} = \left\{ y_1^{(k)}, y_2^{(k)}, \dots, y_{n_k}^{(k)} \right\}, \quad k = 1, 2, \dots, s, \quad \left( \sum_{k=1}^s n_k = n \right) \text{ 其中 } y_j^{(k)} \text{ 是}$$

属于第  $k$  级的数据 ( $1 \leq j \leq n_k$ ). 将  $(x'_{1j}, x'_{2j}, \dots, x'_{mj})$  当作一个样本, 记作  $x_j$ , 即  $x_j \triangleq (x'_{1j}, x'_{2j}, \dots, x'_{mj})$ , 其中  $x'_{ij}$  ( $i = 1, 2, \dots, m$ ) 称为该样本的特性指标. 对应于预报对象  $y$  的等级, 将每个“秩和”因子的数据也相应地分为  $s$  级:

$$x_{i1}^{(k)}, x_{i2}^{(k)}, \dots, x_{im}^{(k)} \quad (i = 1, 2, \dots, m), \quad k = 1, 2, \dots, s, \quad \left( \sum_{k=1}^s m_{(k)} = n \right), \quad i = 1, 2, \dots, m$$

将  $y$  及各“秩和”因子数据 (包括当前值) 中凡属于第  $k$  级的, 均改为“ $k$ ”, 按时间顺序 (样本序号) 将它们列成表 1, 从中去掉那些跟  $y$  的各级对率不高的因子. 不妨假定现在表 1 中的因子跟  $y$  各级的对率都很高 (至少达 80%).

表 1

样本序号	$y$	$y$ 之秩	$y$ 之级	$x'_{11}$ 之级	$x'_{12}$ 之级	...	$x'_{1m}$ 之级
1	$y_1$	$R(y_1)$	$b_1$	$a_{11}$	$a_{21}$	...	$a_{m1}$
2	$y_2$	$R(y_2)$	$b_2$	$a_{12}$	$a_{22}$	...	$a_{m2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n$	$y_n$	$R(y_n)$	$b_n$	$a_{1n}$	$a_{2n}$	...	$a_{mn}$
当前	待报		待报	$a_{10}$	$a_{20}$	...	$a_{m0}$

#### 4. 确定各因子(样本指标)的预报能力

由前面分级可知:“秩和”因子 $x'_i$ 的历史数据中属于第 $k$ 级的个数为 $m_{(k)i}$ ,不属于第 $k$ 级的个数就是 $n - m_{(k)i}$ ;  $y$ 的历史数据中属于第 $k$ 级的个数为 $n_k$ .记

$yx_i(k) = \{b_j = k \text{ 且 } a_{ij} = k, 1 \leq j \leq n\}$  的个数,

$yx_i(k)' = \{b_j \neq k \text{ 且 } a_{ij} \neq k, 1 \leq j \leq n\}$  的个数,

则

$$\frac{yx_i(k)}{m_{(k)i}} = \frac{\{b_j = k \text{ 且 } a_{ij} = k, 1 \leq j \leq n\} \text{ 的个数}}{\{a_{ij} = k, 1 \leq j \leq n\} \text{ 的个数}}$$

是历史数据中 $x'_i$ 属于第 $k$ 级条件下 $y$ 也属于第 $k$ 级的条件频率;

$$\frac{yx_i(k)'}{n - m_{(k)i}} = \frac{\{b_j \neq k \text{ 且 } a_{ij} \neq k, 1 \leq j \leq n\} \text{ 的个数}}{\{a_{ij} \neq k, 1 \leq j \leq n\} \text{ 的个数}}$$

是历史数据中 $x'_i$ 不属于第 $k$ 级情况下 $y$ 也不属于第 $k$ 级的条件频率;

$$\frac{yx_i(k)}{n_k} = \frac{\{b_j = k \text{ 且 } a_{ij} = k, 1 \leq j \leq n\} \text{ 的个数}}{\{b_j = k, 1 \leq j \leq n\} \text{ 的个数}}$$

即历史数据中 $y$ 属于第 $k$ 级条件下 $x'_i$ 也出现在第 $k$ 级的条件频率。因此若记

$$w_i(k) = \frac{yx_i(k)}{m_{(k)i}} \cdot \frac{yx_i(k)'}{n - m_{(k)i}} \cdot \frac{yx_i(k)}{n_k} = \frac{[yx_i(k)]^2 \cdot yx_i(k)'}{m_{(k)i}(n - m_{(k)i})n_k} \quad (1)$$

则 $w_i(k)$ 可以当作衡量“秩和”因子 $x'_i$ 对预报 $y$ 属于第 $k$ 级的能力的指标<sup>[3]</sup>。显然

$$0 \leq w_i(k) \leq 1, \quad k = 1, 2, \dots, s; \quad i = 1, 2, \dots, m$$

$w_i(k)$ 愈大,则根据 $x'_i$ 是否属于第 $k$ 级来报 $y$ 是否属于第 $k$ 级的预报能力就愈强。所谓“预报能力强”其实只是一种模糊概念。设 $Y = \{Y^{(1)}, Y^{(2)}, \dots, Y^{(s)}\}$ 为评语集, $X = \{x'_1, x'_2, \dots, x'_m\}$ 为因子集,由“秩和”因子预报 $y$ 的级别的“预报能力”可用模糊矩阵表示为

$$\tilde{W} = \begin{pmatrix} w_1(1) & w_2(1) & \cdots & w_m(1) \\ w_1(2) & w_2(2) & \cdots & w_m(2) \\ \vdots & \vdots & \ddots & \vdots \\ w_1(s) & w_2(s) & \cdots & w_m(s) \end{pmatrix} \quad \text{记} \quad \begin{pmatrix} \tilde{A}_1 \\ \tilde{A}_2 \\ \vdots \\ \tilde{A}_s \end{pmatrix} \quad (2)$$

其中第 $i$ 行即各个“秩和”因子对预报 $y$ 属于第 $i$ 级的预报能力,它可视为 $X$ 上的模糊子集,记作 $\tilde{A}_i$ 。

#### 5. 确定单因子评判(预报)矩阵

设 $q_{ij}(k)$ 为“秩和”因子 $x'_i$ 属于第 $k$ 级条件下 $y$ 属于第 $j$ 级的样本数,则

$$p_{ij}(k) = \frac{q_{ij}(k)}{m_{(k)i}} \quad (3)$$

是在 $x'_i$ 属于第 $k$ 级条件下, $y$ 属于第 $j$ 级的条件频率。从表1中统计出“秩和”因子 $x'_i$ 属于第 $k$ 级条件下 $y$ 属于第 $j$ 级的样本数( $j = 1, 2, \dots, s$ ),便可建立该“秩和”因子和预报对象级别之间的模糊关系矩阵

$$\tilde{N}(x'_i, y) = \begin{bmatrix} p_{i1}(1) & p_{i2}(1) & \cdots & p_{is}(1) \\ p_{i1}(2) & p_{i2}(2) & \cdots & p_{is}(2) \\ \vdots & \vdots & \ddots & \vdots \\ p_{i1}(s) & p_{i2}(s) & \cdots & p_{is}(s) \end{bmatrix}, \quad i = 1, 2, \dots, m \quad (4)$$

对于任一样本 $X_l$ , 根据其特性指标级别 $a_{1l}, a_{2l}, \dots, a_{ml}$ , 分别从 $\tilde{N}(x'_1, y)$ 中取出第 $a_{1l}$ 行, 从 $\tilde{N}(x'_2, y)$ 中取出第 $a_{2l}$ 行,  $\dots$ , 从 $\tilde{N}(x'_m, y)$ 中取出第 $a_{ml}$ 行, 将它们组成一模糊矩阵 $\tilde{R}(a_{1l}, a_{2l}, \dots, a_{ml})$ :

$$\tilde{R}(a_{1l}, a_{2l}, \dots, a_{ml}) = \begin{bmatrix} p_{11}(a_{1l}) & p_{12}(a_{1l}) & \cdots & p_{1s}(a_{1l}) \\ p_{21}(a_{2l}) & p_{22}(a_{2l}) & \cdots & p_{2s}(a_{2l}) \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1}(a_{ml}) & p_{m2}(a_{ml}) & \cdots & p_{ms}(a_{ml}) \end{bmatrix}$$

记 $\begin{bmatrix} \tilde{B}_{1l}, \tilde{B}_{2l}, \dots, \tilde{B}_{sl} \end{bmatrix}$  (5)

其中 $\tilde{B}_{jl} = [p_{1j}(a_{1l}), p_{2j}(a_{2l}), \dots, p_{mj}(a_{ml})]^T$ 为列向量( $j = 1, 2, \dots, s$ ). 由“秩和”因子 $x'_i$ 所属级别 $a_{il}$ 来评判(预报) $y$ 的级别, 其结果可视为评语集 $Y$ 上的模糊子集:

$$\tilde{R}_{il} = [p_{i1}(a_{il}), p_{i2}(a_{il}), \dots, p_{is}(a_{il})]$$

$\tilde{R}_{il}$ 是 $\tilde{R}(a_{1l}, a_{2l}, \dots, a_{ml})$ 中的第 $i$ 行, 故 $\tilde{R}(a_{1l}, a_{2l}, \dots, a_{ml})$ 即单因子评判(预报)矩阵,  $\tilde{R}(a_{1l}, a_{2l}, \dots, a_{ml})$ 是由 $X$ 到 $Y$ 的一个模糊关系.

## 6. 用模糊综合评判的方法评定(预报) $y$ 的级别

当 $l=0$ , 上述式(5)给出的即是在各因子当前所属级别条件下的单因子评判矩阵,  $\tilde{B}_{j0}$ 给出在各因子当前所属级别条件下 $y$ 属于第 $j$ 级的条件频率. 注意到因子对 $y$ 的不同级别的预报能力是不同的, 故对 $y$ 的不同级别取不同的权向量跟单因子评判矩阵的相应列作合成运算(合成运算采用普通的“加”、“乘”算子: “+”和“.”), 以求得模糊综合评定(预报)结果. 详细点说就是: 取 $\tilde{A}_j$ 与 $\tilde{B}_{j0}$ 作合成运算得

$$\tilde{C}_j = \tilde{A}_j \tilde{B}_{j0} = w_1(j)p_{1j}(a_{10}) + w_2(j)p_{2j}(a_{20}) + \cdots + w_m(j)p_{mj}(a_{m0}), \quad j = 1, 2, \dots, s \quad (6)$$

作归一化处理, 令

$$c^*_j = \frac{\tilde{C}_j}{\sum_{k=1}^s \tilde{C}_k}, \quad j = 1, 2, \dots, s \quad (7)$$

综合评判(预报)结果是

$$\tilde{C} = (C^*_1, C^*_2, \dots, C^*_s) \quad (8)$$

根据最大从属原则<sup>1)</sup>, 若

1) 其实归一化处理只是为了形成从属函数.

$$C^*_{\cdot h} = \bigvee_{j=1}^s C^*_{\cdot j} \quad (9)$$

则预报  $y$  属于第  $h$  级。

### 7. 回报检验

类似于上一步骤, 由  $\tilde{W}$  和  $\tilde{R}(a_{1l}, a_{2l}, \dots, a_{ml})$  利用模糊综合评判的方法即可求得第  $l$  个样本的回报级别 ( $l=1, 2, \dots, n$ )。若回报准确率低于80%, 则须适当改换预报因子, 再重新计算。

### 8. 用模糊模式识别方法预报 $y$ 的数值

假如回报准确率高 (达80%以上), 并已预报出  $y$  属于第  $h$  级, 则将  $y$  的历史资料中凡属于第  $h$  级的数据及其所对应的各“秩和”因子数据列入表2。

设  $y$  与各“秩和”因子的相关系数分别为  $r_{iy}$ ,  $i=1, 2, \dots, m$ 。可用  $|r_{iy}|$  作为衡量  $x'_i$  对预报  $y$  之值的作用大小的指标。令

$$e_i = \frac{|r_{iy}|}{\sum_{k=1}^m |r_{ky}|}, \quad i=1, 2, \dots, m \quad (10)$$

显然  $\sum_{i=1}^m e_i = 1$ ,  $0 < e_i \leq 1 (i=1, 2, \dots, m)$ 。将  $e_i$  列入表2。

表 2

样本序号	$y^{(h)}$	$x'_1$	$x'_2$	...	$x'_m$	$X_0$ 的从属函数值
$h(1)$	$y_1^{(h)}$	$x'_{1h(1)}$	$x'_{2h(1)}$	...	$x'_{mh(1)}$	$\mu_{\tilde{x}_1^{(h)}}(X_0)$
$h(2)$	$y_2^{(h)}$	$x'_{1h(2)}$	$x'_{2h(2)}$	...	$x'_{mh(2)}$	$\mu_{\tilde{x}_2^{(h)}}(X_0)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$h(n_h)$	$y_{n_h}^{(h)}$	$x'_{1h(n_h)}$	$x'_{2h(n_h)}$	...	$x'_{mh(n_h)}$	$\mu_{\tilde{x}_{n_h}^{(h)}}(X_0)$
当前值		$x'_{10}$	$x'_{20}$	...	$x'_{m0}$	
权数		$e_1$	$e_2$	...	$e_m$	

记  $m$  维向量空间为  $\mathcal{X}$ , 则任一样本  $X = (x'_1, x'_2, \dots, x'_m)$  均为  $\mathcal{X}$  中的元素, 称  $\mathcal{X}$  为样本空间。由诸“秩和”因子构成的样本跟  $y$  之值的对应关系实际上只是一种模糊关系。对于表2中的任一  $y^{(h)}$ , 能预报它的样本应视为  $\mathcal{X}$  上的一个模糊子集, 记作  $\tilde{x}_j^{(h)}$ , 其从属函数可规定如下:

$$\mu_{\tilde{x}_j^{(h)}}(x) = \frac{\sum_{k=1}^m e_k x'_k x'_{kh(j)}}{\sqrt{\sum_{k=1}^m [x'_k]^2 \cdot \sum_{k=1}^m [x'_{kh(j)}]^2}}, \quad X = (x'_1, x'_2, \dots, x'_m) \in \mathcal{X} \quad (11)$$

取  $\tilde{y} = \bigcup_{j=1}^h X_j^{(h)}$ , 其从属函数为

$$\mu_{\tilde{y}}(X) = \bigvee_{j=1}^h \mu_{X_j^{(h)}}(X), \quad X \in \mathcal{X} \quad (12)$$

如果

$$\mu_{X_{j_0}^{(h)}}(X_0) = \max_{1 \leq j \leq n_h} \mu_{X_j^{(h)}}(X_0) = \mu_{\tilde{y}}(X_0) \quad (13)$$

则可预报  $\hat{y}_{n+1} \approx y_{j_0}$ .

### 三、预报实例及效果比较

根据福建省泉州市气象台1959—1984年实测资料, 预报福建省泉州市1985年3月上旬雨量。

预报对象用  $y$  表示。利用秩相关检验法从631个气象因子中选出15个跟  $y$  关系较为密切的因子, 其实测数据见表3 (因子观测时间均比  $y$  提前)。对这些因子作秩变换并用试探组合的办法将5个因子组成一组, 取其秩数之和作为一个新因子,  $y$  及各新因子的数据列入表4。新因子的组成情况及其与  $y$  的秩相关系数均列入表5 (括弧里的内容为构成新因子的原因子代号及其与  $y$  的相关系数)。从表5可见新因子与  $y$  的秩相关系数一般均显著地高于原因子跟  $y$  的相关系数。

将  $y$  的历史数据分为三级, 分级标准是:

第一级  $y_j \leq 3.3\text{mm}$ ;

第二级  $3.3\text{mm} < y_j \leq 40\text{mm}$ ;

第三级  $y_j > 40\text{mm}$ 。

将各“秩和”因子也分成三级, 分级标准是:

第一级  $x'_1 \leq 63$ ,  $x'_2 \leq 60$ ,  $x'_3 \leq 60$ ;

第二级  $63 < x'_1 \leq 90$ ,  $60 < x'_2 \leq 88$ ,  $60 < x'_3 \leq 88$ ;

第三级  $x'_1 > 90$ ,  $x'_2 > 88$ ,  $x'_3 > 88$ 。

将  $y$  及各“秩和”因子数据中凡属于第  $k$  级的, 均改为“ $k$ ”, 按时间顺序列表6。根据表6统计出  $n_k$ ,  $yx_i(k)$ ,  $yx_i(k)'$ ,  $m_{(k)i}$ ,  $(n - m_{(k)i})$  ( $i = 1, 2, 3$ ;  $k = 1, 2, 3$ ), 将它们列入表7。根据表7按式(1)计算各“秩和”因子预报  $y$  属于各级的能力  $w_i(k)$ , 按式(2)写成模糊矩阵

$$\tilde{W} = \begin{bmatrix} w_1(1) & w_2(1) & w_3(1) \\ w_1(2) & w_2(2) & w_3(2) \\ w_1(3) & w_2(3) & w_3(3) \end{bmatrix} = \begin{bmatrix} 0.8182 & 0.9 & 0.6003 \\ 0.5547 & 0.4987 & 0.4424 \\ 0.6032 & 0.3214 & 0.7917 \end{bmatrix} \text{ 记 } \begin{bmatrix} \tilde{A}_1 \\ \tilde{A}_2 \\ \tilde{A}_3 \end{bmatrix}$$

表 3

年份	X <sub>73</sub>	X <sub>100</sub>	X <sub>108</sub>	X <sub>171</sub>	X <sub>204</sub>	X <sub>207</sub>	X <sub>282</sub>	X <sub>340</sub>	X <sub>342</sub>	X <sub>420</sub>	X <sub>497</sub>	X <sub>498</sub>	X <sub>532</sub>	X <sub>547</sub>	X <sub>631</sub>
1960	66.4	66	77	1.0	132.4	108.5	75	0.77	0.58	488	-6	10	228	-3	4907
1961	1.8	12	23	7.5	47.1	3.4	11	0.49	0.65	493	-6	20	243	-7	3559
1962	97.5	78	58	1.0	9.6	219.7	72	0.66	0.89	492	-12	4	242	11	7194
1963	88.4	104	66	3.7	101.4	9.8	100	0.68	0.65	496	-13	5	242	20	2522
1964	403.9	98	71	1.1	79.5	299.3	100	0.61	0.69	484	-19	-4	241	3	4418
1965	0.0	8	54	5.2	116.4	8.5	12	0.81	0.54	496	-18	-8	255	-3	2606
1966	205.7	84	27	4.5	51.3	127.5	78	0.83	0.61	494	-2	16	240	5	2821
1967	29.8	38	29	6.1	67.3	79.5	30	0.64	0.62	497	-4	9	241	5	2540
1968	71.9	31	41	3.0	31.9	34.1	31	0.67	0.64	489	-21	-7	240	37	2015
1969	19.8	23	3	5.3	90.9	0.0	32	0.57	0.62	502	13	20	248	1	1732
1970	17.2	16	36	6.1	99.6	6.1	22	0.68	0.49	503	4	17	254	3	2884
1971	54.4	44	34	3.1	32.6	96.5	36	0.64	0.80	496	3	25	250	25	4423
1972	102.5	42	83	4.9	29.9	29.4	47	0.68	0.64	500	-9	7	242	-9	2741
1973	181.0	49	64	2.9	0.7	127.1	38	0.64	0.70	492	-26	-17	240	23	5477
1974	368.0	48	56	0.2	56.8	95.5	33	0.72	0.68	494	-14	2	246	32	4543
1975	12.0	17	57	3.7	93.3	1.5	24	0.60	0.60	506	-2	19	242	2	2092
1976	18.3	51	49	7.3	83.5	122.5	60	0.74	0.57	505	-10	8	257	18	4057
1977	60.7	48	17	5.1	44.0	9.7	40.4	0.86	0.68	490	-17	-5	238	27	3793
1978	8.1	34	58	5.0	177.2	6.5	26.4	0.59	0.66	500	7	25	254	-14	2366
1979	4.9	46	7	4.1	128.1	9.6	38.6	0.74	0.65	497	-6	19	250	14	2655
1980	2.7	39	15	5.2	91.5	1.5	34	0.61	0.53	488	-16	-10	247	11	2500
1981	103.5	42	53	5.5	17.3	73.9	40	0.72	0.64	505	6	33	243	-10	3371
1982	284.6	15	28	2.8	53.4	80.9	20	0.62	0.45	498	-6	18	245	-26	3433
1983	164.5	63	38	4.8	47.9	0.2	41	0.67	0.55	498	-5	34	245	22	2488
1984	4.9	45	10	5.6	0.0	10.4	38	0.80	0.71	496	-16	6	251	19	2345
当前值															
1985	5.9	10	29	2.3	51.8	158.0	6	0.68	0.70	494	-23	-10	242	20	4117

表 4

年份	y	x' <sub>1</sub>	x' <sub>2</sub>	x' <sub>3</sub>	年份	y	x' <sub>1</sub>	x' <sub>2</sub>	x' <sub>3</sub>
1960	0.4	29.5	47.5	58.0	1961	16.9	81.0	99.5	88.0
1962	1.9	26.5	26.0	43.0	1963	0.7	52.5	45.0	50.5
1964	7.7	18.5	30.5	43.0	1965	29.5	87.0	76.5	82.0
1966	1.9	42.5	60.0	46.0	1967	35.7	68.5	85.5	76.5
1968	0.2	61.0	48.0	41.5	1969	117.8	113.0	98.5	94.5

表 4 (续)

年份	y	x' <sub>1</sub>	x' <sub>2</sub>	x' <sub>3</sub>	年份	y	x' <sub>1</sub>	x' <sub>2</sub>	x' <sub>3</sub>
1970	44.2	98.5	94.5	90.5	1971	7.1	65.5	66.5	49.5
1972	0.0	63.0	38.0	74.0	1973	0.0	19.0	21.5	37.5
1974	0.4	45.5	47.0	28.5	1975	42.7	95.5	76.0	101.0
1976	3.4	55.0	70.0	76.0	1977	0.3	40.5	59.0	28.5
1978	60.0	104.5	89.0	94.5	1979	19.9	77.0	87.5	63.5
1980	98.2	81.0	74.5	80.5	1981	16.4	72.0	69.0	75.5
1982	64.9	71.0	76.0	89.5	1983	30.6	81.5	72.0	64.0
1984	10.1	75.5	67.5	49.0	1985 (待报)	46.0	61.0	51.5	

表 5

新因子构成情况	新因子跟y的相关系数
x' <sub>1</sub> (x <sub>100</sub> , x <sub>267</sub> , x <sub>497</sub> , x <sub>532</sub> , x <sub>831</sub> )	0.8046 (-0.5598, -0.5954, 0.5425, 0.5592, -0.4906)
x' <sub>2</sub> (x <sub>292</sub> , x <sub>498</sub> , x <sub>108</sub> , x <sub>284</sub> , x <sub>171</sub> )	0.8265 (-0.5508, 0.4692, -0.4725, 0.4593, 0.4533)
x' <sub>3</sub> (x <sub>342</sub> , x <sub>340</sub> , x <sub>547</sub> , x <sub>73</sub> , x <sub>420</sub> )	0.8135 (-0.4911, -0.4804, -0.4769, -0.4542, 0.4500)

表 6

年份	y之秩	y之级	x' <sub>1</sub> 之级	x' <sub>2</sub> 之级	x' <sub>3</sub> 之级	回报y之秩	回报对否
1960	5.5	1	1	1	1	1	✓
1961	15	2	2	3	2	2	✓
1962	3.5	1	1	1	1	1	✓
1963	7	1	1	1	1	1	✓
1964	12	2	1	1	1	1	×
1965	17	2	2	2	2	2	✓
1966	8.5	1	1	1	1	1	✓
1967	19	2	2	2	2	2	✓
1968	3	1	1	1	1	1	✓
1969	25	3	3	3	3	3	✓
1970	21	3	3	3	3	3	✓
1971	11	2	2	2	1	2	✓
1972	1.5	1	1	1	2	1	✓
1973	1.5	1	1	1	1	1	✓
1974	5.5	1	1	1	1	1	✓
1975	20	3	3	2	3	3	✓
1976	10	2	1	2	2	2	✓
1977	4	1	1	1	1	1	✓



表 6 (续)

年份	y 之秩	y 之级	$x'_1$ 之级	$x'_2$ 之级	$x'_3$ 之级	回报y之级	回报对否
1978	22	3	3	3	3	3	✓
1979	16	2	2	2	2	2	✓
1980	24	3	2	2	2	2	×
1981	14	2	2	2	2	2	✓
1982	23	3	2	2	3	3	✓
1983	18	2	2	2	2	2	✓
1984	13	2	2	2	1	2	✓
当前		(实际)				(预报)	回报率
1985		1	1	1	1	1	92%

表 7

		y	$x'_1$	$x'_2$	$x'_3$
第一级	$n_1$	9			
	$yx_i(1)$		9	9	8
	$yx_i(1)'$		14	15	13
	$m_{(1)i}$		11	10	11
	$n - m_{(1)i}$		14	15	14
第二级	$n_2$	10			
	$yx_i(2)$		8	8	7
	$yx_i(2)'$		13	12	13
	$m_{(2)i}$		10	11	9
	$n - m_{(2)i}$		15	14	16
第三级	$n_3$	6			
	$yx_i(3)$		4	3	5
	$yx_i(3)'$		19	18	19
	$m_{(3)i}$		4	4	5
	$n - m_{(3)i}$		21	21	20

根据表 6 的数据, 按式(3)、(4)建立每个“秩和”因子跟 y 各级之间的模糊关系矩阵

$$N(x'_1, y) = \begin{bmatrix} \frac{9}{11} & \frac{2}{11} & \frac{0}{11} \\ \frac{0}{10} & \frac{8}{10} & \frac{2}{10} \\ \frac{0}{4} & \frac{0}{4} & \frac{4}{4} \end{bmatrix} = \begin{bmatrix} 0.8182 & 0.1818 & 0 \\ 0 & 0.8 & 0.2 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\tilde{N}(x'_2, y) = \begin{bmatrix} \frac{9}{10} & \frac{1}{10} & 0 \\ 0 & \frac{8}{11} & \frac{3}{11} \\ \frac{0}{4} & \frac{1}{4} & \frac{3}{4} \end{bmatrix} = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0 & 0.7273 & 0.2727 \\ 0 & 0.25 & 0.75 \end{bmatrix}$$

$$\tilde{N}(x'_3, y) = \begin{bmatrix} \frac{8}{11} & \frac{3}{11} & \frac{0}{11} \\ \frac{1}{9} & \frac{7}{9} & \frac{1}{9} \\ \frac{0}{5} & \frac{0}{5} & \frac{5}{5} \end{bmatrix} = \begin{bmatrix} 0.7273 & 0.2727 & 0 \\ 0.1111 & 0.7778 & 0.1111 \\ 0 & 0 & 1 \end{bmatrix}$$

已知各“秩和”因子的当前值的级别  $a_{10} = a_{20} = a_{30} = 1$ , 故从上述三个矩阵中各取出一行组成一模糊矩阵  $\tilde{R}(1, 1, 1)$

$$\tilde{R}(1, 1, 1) = \begin{bmatrix} 0.8182 & 0.1818 & 0 \\ 0.9 & 0.1 & 0 \\ 0.7273 & 0.2727 & 0 \end{bmatrix} \text{ 记 } [\tilde{B}_{10}, \tilde{B}_{20}, \tilde{B}_{30}]$$

由  $\tilde{R}(1, 1, 1)$  及  $w$ , 按式(6)求合成运算即得

$$\tilde{C} = (1.9160493, 0.2713568, 0) \triangle (c_1, c_2, c_3)$$

(未作归一化处理)。由于  $c_1 = \sum_{j=1}^3 c_j$ , 故预报  $y$  将属于第一级, 即预报泉州市1985年3月上旬雨量小于3.3mm, 与实况(3.2mm)相符。

回报检验:

从表6可知1960、1962—1964、1966、1968、1973、1974、1977等年份各“秩和”因子的级别也都是1, 其单因子评判矩阵都是  $\tilde{R}(1, 1, 1)$ , 所以回报  $y$  的级别也都是1。1961年各“秩和”因子的级别分别是2, 3, 2, 按式(5)其单因子评判矩阵应取  $\tilde{R}(2, 3, 2)$

$$\tilde{R}(2, 3, 2) = \begin{bmatrix} 0 & 0.8 & 0.2 \\ 0 & 0.25 & 0.75 \\ 0.1111 & 0.7778 & 0.1111 \end{bmatrix}$$

按式(6)求合成运算得

$$\tilde{C} = (0.0666933, 0.9125337, 0.4496478) \triangle (c_1, c_2, c_3)$$

由于  $c_2$  最大, 故回报1961年  $y$  属于第二级。如此逐年回报, 将所有回报结果列入表6, 跟  $y$  的实际级别对比, 回报准确率达92%。显然本文所述方法比“跟踪分级预报法”<sup>[3]</sup>效果更好(根据同样的数据, 按文[3]的方法回报准确率为88%)。

作  $y$  的数值预报。

前面已经预报出1985年  $y$  将属于第一级, 于是将  $y$  的历史资料中凡属于第一级的数据及相应的各“秩和”因子数据列出如表 8。根据表 5 的相关系数按式(10)计算得

$$c_1 = \frac{0.8046}{2.4446} = 0.3291, \quad c_2 = \frac{0.8265}{2.4446} = 0.3381, \quad c_3 = \frac{0.8135}{2.4446} = 0.3328$$

将它们填入表 8 最末一行。然后按式(11)计算诸从属函数值  $\mu_{\tilde{x}_j^{(1)}}(X_0) (j=1, 2, \dots, 9)$ , 将

表 8

年份	$y^{(1)}$	$x'_1$	$x'_2$	$x'_3$	$\mu_{\tilde{x}_j^{(1)}}(X_0)$
1960	0.4	29.5	47.5	58.0	0.3260593
1962	1.9	26.5	26.0	43.0	0.3198927
1963	0.7	52.5	45.0	50.5	0.3280943
1966	1.9	42.5	60.0	46.0	0.334025
1968	0.2	61.0	48.0	41.5	0.3236452
1972	0.0	63.0	38.0	74.0	0.3126251
1973	0.0	19.0	21.5	37.5	0.3157738
1974	0.4	45.5	47.0	28.5	0.3265087
1977	0.3	40.5	59.0	28.5	0.326662
$x'_{i0}$		46.0	61.0	51.5	
$c_i$		0.3291	0.3381	0.3328	

它们填入表 8 最右一列, 由此可见

$$\mu_{\tilde{x}_4^{(1)}}(X_0) = \max_{1 \leq j \leq 9} \mu_{\tilde{x}_j^{(1)}}(X_0)$$

故预报  $\hat{y}_{85} \approx y_{86} = 1.9\text{mm}$ , 即泉州市1985年 3 月上旬降雨量约在 1.9mm 左右, 与实况 (3.2mm) 很接近。

## 参 考 文 献

- [1] 丁士晟, 多元分析方法及其应用, 吉林人民出版社, (1981).
- [2] 汪培庄, 模糊集合论及其应用, 上海科学技术出版社, (1983).
- [3] 吴绍敏、彭沛, 以秩和为因子的跟踪预测方法, 数理统计及应用概率, 2(1987).

## A Fuzzy Multiple-Order Discriminatory Analysis and Prediction with Combined Sample Indexes

Chen Zhidiao

### Abstract

This paper tries to combine fuzzy mathematic method with multi-analysis. The method in question would make full use of the information of sample indexes, both linear and nonlinear. It is capable of treating fuzzy information free from the restriction of probability distribution of sampled data. It would be able to forecast the trend and definite value of the object.

In a word, it improves and expands the extent of mutipleorder discriminatory analysis and prediction.