文章编号: 1000-6788(2000) 09-0123-02

灰色系统分析中存在的两个基本问题

刘震宇
(厦门大学自动化系,福建 厦门 361005)

摘要: 主要讨论灰色系统关联分析中关联系数计算公式存在的问题和 GM(1, N)模型的系数辨识中存在的独立变量之间的共线性问题.

关键词: 关联系数;共线性
中图分类号: TP393.2; TP393.3; TP311.5 ⒇

Two Problems Existing in Gray Systems Analysis

LIU Zhen-yu
(Dept. of Automation, Xiamen University, Xiamen 361005)

Abstract: In this paper, the problem for computing the associate coefficient in gray systems analysis is discussed. The correlation between independent variables in recognizing the coefficients of GM(1, N) model is suggested.

Keywords: associate coefficient; correlation

灰色系统关联分析和模型参数辨识中存在两个基本问题,其一是按照关联系数计算公式计算出的值存在不合理的下限,其二是 GM(1, N) ($N \geq 2$)模型的独立变量之间存在共线性问题. 这在实际应用中必须引起重视.

灰色系统关联分析中的关联系数计算公式存在的问题在有关文献中被讨论过多次,但是按照关联系数计算公式计算出的值,存在不合理下限的问题却被忽略了. 例如,当分辨系数 $d = 0.5$时,无论参考数列和对比数列为何,相应的关联系数的下限值均为 0.3333,这显然是不合理的.

在应用数理统计方法辨识 GM(1, N)模型的参数时,由于模型中的独立变量之间存在共线性关系,这使 GM(1, N)模型不合适于作预测,因为共线性问题会引起较大的误差. 而 GM(1, 1)模型中不存在这个问题.

下面,我们就以上两个问题作进一步的讨论.

1 灰色关联系数计算公式存在的问题

按参考文献[1],灰色关联系数计算公式的形式如下:

$$a_i = \frac{\min\limits_i \min\limits_k |x_0(k) - x_i(k)| + d \max\limits_i \max\limits_k |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + d \max\limits_i \max\limits_k |x_0(k) - x_i(k)|}$$

其中, $x_0(k)$为参考数列的第 k 个数值, $x_i(k)$为比较数列 i 的第 k 个数值, d为分辨系数, d在 0和 1之间取值.

考虑到不同数据的量纲不同,通常对各数列进行初始化处理,即用每一个数列的第一个数 $x_i(1)$除其它数 $x_i(k)$,这样可以使数列无量纲化,又可得到公共交叉点. 因为初始化处理,使 $\min\limits_i \min\limits_k |x_0(k) - x_i(k)| = 0$,故上式可改写为

⒇ 收稿日期: 1999-02-26
资助项目: 福建省留学回国人员基金(K36001)

$$a_i = \frac{d \max\limits_i \max\limits_k |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + d \max\limits_i \max\limits_k |x_0(k) - x_i(k)|}$$

一般地, d取 0.5,则可看到如下情形:① 当 $|x_0(k) - x_i(k)| = \max\limits_i \max\limits_k |x_0(k) - x_i(k)|$ 时, $a_i(k) = \frac{d}{1+d} = \frac{0.5}{1.5} = 0.3333$; ② 当 $|x_0(k) - x_i(k)| = \min\limits_i \min\limits_k |x_0(k) - x_i(k)| = 0$ 时, $a_i(k) = \frac{d}{d} = \frac{0.5}{0.5} = 1$. 于是, $a_i(k) \in [0.333, 1]$. 类似地,若 d取 1,则 $a_i(k) \in [0.5, 1]$.

按 d取不同的值,可得出下列表:

d	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$\min\{a_i(k)\}$	0.091	0.167	0.231	0.286	0.333	0.375	0.412	0.444	0.474	0.5

显然,按关联度的一般表达式 $V_i = \frac{1}{N}\sum\limits_{k=1}^{N} a_i(k)$ 计算,当 d取确定的值时, V_i 的值不会小于 $\min\{a_i(k)\}$. 因此,关联度大小无绝对意义,只有相对的意义,特别是 d取较大的值时更是如此. 也就是说,只有在考察多个比较数列与参考数列进行比较时,要确定比较数列与参考数列关联的相对程度,则关联度越大的,与参考数列越有紧密关系. 单独考察某一数列与参考数列关联度的大小,无法确定该数列与参考数的关系的紧密程度.

从以上讨论可知, d取得越小的值时, V_i 值的绝对意义就越大,因为 $\min\{a_i(k)\}$的值就越趋近于零.

实际上,存在着许多关联度的分析方法,有兴趣的读者可参考文献[2]中的有关内容.

2 GM(1,N) (N≥2)模型中独立变量之间的共线性问题

按数理统计的方法,辨识 GM(1,N)模型中的参数时,常常发现各个独立变量之间存在着较强的共线性问题. 而共线性的存在将会引入误差,从而使模型的精度变差.

一般地, GM(1,N)模型的形式如下所示:

$$(1 + 0.5a)x_1^{(0)}(k) + ax_1^{(1)}(k-1) = \sum_{i=2}^{N} b_i x_i^{(1)}(k)$$

它可改写为 $x_1^{(0)}(k) = \frac{1}{1+0.5a}\left[-ax_1^{(1)}(k-1) + \sum\limits_{i=2}^{N} b_i x_i^{(1)}(k)\right]$

其中, $x_1^{(0)}(k)$为因变量,即系统的主行为变量, $x_i^{(1)}(k)$和 $x_1^{(1)}(k-1)$为自变量,即系统的作用变量.

由于 $x_1^{(1)}(k-1)$和 $x_i^{(1)}(k)$之间常有共线性关系,无法通过数理统计检验,因此影响模型的精度,无法用于正确的预测. 这也是 GM(1,N)模型的预测精度比 GM(1,1)模型差的原因.

事实上, $x_1^{(0)}(k) = x_1^{(1)}(k) - x_1^{(1)}(k-1)$,若 $x_1^{(0)}(k)$和 $x_1^{(1)}(k-1)$之间有线性关系,而 $x_i^{(1)}(k)$和 $x_1^{(0)}(k)$有线性关系,则 $x_i^{(1)}(k)$和 $x_1^{(1)}(k-1)$之间也常有线性关系. 因为,当对所有的初始数据作规范化后,如果 $x_i^{(1)}(k)$和 $x_1^{(1)}(k-1)$均服从灰指数规律,它们之间存在着一定的相似性,因此, GM(1,N)模型中独立变量之间常有共线性关系. 而这种情况常无法用常规的方法予处理. 根据我们在许多项目中应用 GM(1,N)模型的经验表明,这种情况是经常存在的.

3 结论

从以上分析可知: 1) 按灰色关联度计算得出的关联度只有相对比较时才有意义,无绝对意义. 2) 按 GM(1,N)模型无法作正确的预测,因为模型的自变量之间常存在共线性关系,即使应用 GM(1,2)模型也必须对有关变量之间的关系进行检验,以防共线性引起的误差.

参考文献:

[1] 邓聚龙. 灰色系统预测与决策[M]. 武汉:华中理工大学出版社, 1986.

[2] 贺建勋. 建模与数学模型[M]. 福州:福建科学技术出版社, 1995.