

基于数值积分的 GM(1,1) 模型的应用

江银¹
¹(浙江工业大学)

摘 要 文中分析了 GM(1,1) 模型及其背景值, 并从梯形公式到插值公式和数值积分入手展开了延申。梯形公式较为简单, 且也能良好地对 GM(1,1) 模型进行估计和预测; 后面探讨的插值公式和数值积分有一定的价值, 但不具有普适性, 且具有插值次数过高和计算冗杂等缺点。

关键词 灰色理论 GM(1,1) 模型 Newton-Cotes 公式 插值公式

灰色理论认为一切随机变量都是在一定范围内、一定时间段上变化的灰色量及灰色过程。数据处理不去寻找其统计规律和概率分布, 而是对原始数据作一定处理后, 使其成为有规律的时间序列数据, 在此基础上建立数学模型。

GM 系列模型是灰色预测理论的基本模型, 尤其是 GM(1,1) 模型, 应用十分广泛。灰色 GM(1,1) 模型对高增长指数序列拟合常常产生滞后误差, GM(1,1) 模型中背景值构造方法是影响其精度和适应性的重要因素。本文将来讨论背景值两种构造方法的简易性、普适性以及模型预测的准确性。

1 GM(1,1) 预测模型的建模机理

一次累加 GM(1,1) 模型是最常用的一种灰色动态预测模型, 该模型由一个单变量的一阶微分方程构成。其建模过程如下:

设原始非负数据序列为:

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

其中 $x^{(0)}(i) > 0, i = 1, 2, \dots, n$ 。

对原始数据作一次累加:

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

其中 $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i), i = 1, 2, \dots, n$ 。由一阶生成模块 $X^{(1)}$ 建立模型 GM(1,1), 对应的微分方程为:

$$\frac{dx^{(1)}(t)}{dt} + ax^{(1)}(t) = u \quad (3)$$

其中 a 和 u 为待辨识常数。

待辨识常数的最小二乘解为:

$$\hat{a} = [\hat{a}, \hat{u}] = (B^T B)^{-1} B^T Y \quad (4)$$

其中 $Y = [x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)]^T, B =$

$$\begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}, \text{背景值 } z^{(1)}(k+1) = \frac{1}{2}[x^{(1)}(k+1) + x^{(1)}(k)].$$

方程 (3) 的离散解为:

$$\hat{x}^{(1)}(k+1) = [x^{(1)}(1) - \frac{u}{\hat{a}}]e^{-\hat{a}k} + \frac{u}{\hat{a}} \quad (5)$$

还原到原始数据为:

$$\begin{aligned} \hat{x}^{(0)}(k+1) &= \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \\ &= (1 - e^{\hat{a}})[x^{(1)}(1) - \frac{u}{\hat{a}}]e^{-\hat{a}k} \end{aligned} \quad (6)$$

从公式 (4) 可以看出拟合和预测的精度取决于待辨识常数 a 和 u , 而 a 和 u 的求解依赖于背景值 $z^{(1)}(k+1)$, 它会直接影响所构建的模型的精度和适应性。

2 对于 GM(1,1) 模型背景值改进的讨论

上述背景值的求法采用了数值积分中的梯形法, 因为我们知道两个相邻点的信息, 可以近似成直线来求取它们围成区域的面积。

然而不管是牛顿插值还是拉格朗日插值, 我们无法只通过相邻两个点来近似成抛物线或更高次数曲线的情况。原论文采用了将 $X^{(1)}$ 的所有点进行牛顿插值后, 将每两个节点之间的四等分点的值和它们本身的值来近似得到次数更高的更精确的数值积分。

原论文将 $X^{(1)}$ 都参与了插值, 构成了一个高次插值多项式, 同时也应注意由于次数过高而产生的龙格现象。因此该方法并不具有普适性。

2.1 插值公式和 Newton-Cotes 公式

根据数值计算的知识我们可以得到 n 个点的插值多项式可以用以下公式得到:

$$\begin{aligned} N_n(x) = & f(x_0) + f[x_0, x_1](x - x_0) \\ & + f[x_0, x_1, x_2](x - x_0)(x - x_1) \\ & + \cdots \\ & + f[x_0, x_1, \cdots, x_n](x - x_0) \cdots (x - x_{n-1}) \end{aligned} \quad (7)$$

将积分区间 $[a, b]$ 划分为 n 等分, 步长 $h = \frac{b-a}{n}$, 选取等距节点 $x_k = a + kh$ 构造出的插值型积分公式

$$I_n = (b-a) \sum_{k=0}^n C_k^{(n)} f(x_k) \quad (8)$$

称为 Newton-Cotes 公式, 其中 $C_k^{(n)}$ 称为 Cotes 系数。

$$\begin{aligned} C_k^{(n)} = & \frac{h}{b-a} \int_0^n \prod_{j=0, j \neq k}^n \frac{t-j}{k-j} dt \\ = & \frac{(-1)^{n-k}}{nk!(n-k)!} \int_0^n \prod_{j=0, j \neq k}^n (t-j) dt \end{aligned} \quad (9)$$

当 $n=1$ 时, 求积公式就是我们熟悉的梯形公式。当 $n=2$ 时, 求积公式

$$S = \frac{b-a}{6} [f(a) + 4f(\frac{a+b}{2}) + f(b)] \quad (10)$$

就是 Simpson 公式。当 $n=4$ 时, 求积公式

$$C = \frac{b-a}{90} [7f(x_0) + 32f(x_1) + 12f(x_2) + 32f(x_3) + 7f(x_4)] \quad (11)$$

就是 Cotes 公式, 这里 $x_k = a + kh, h = \frac{b-a}{4}$ 。

2.2 基于数值积分公式的背景值改进

如下是基于 Cotes 公式的改进方法。首先取 (2) 中的 $X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(n)\}$ 。令 $Y(k) = k, k=1, 2, \cdots, n$ 。把 $[(k), x^{(1)}(k)], k=1, 2, \cdots, n$ 作为对应曲线上的等间距点的坐标, 用牛顿插值求出 $Y(k+0.25), Y(k+0.5), Y(k+0.75), k=1, 2, \cdots, n$ 对应的横纵坐标。此时背景值 $z^{(1)}(k+1) = \frac{1}{90} [7x^{(1)}(k) + 32x^{(1)}(k+0.25) + 12x^{(1)}(k+0.5) + 32x^{(1)}(k+0.75) + 7x^{(1)}(k+1)]$, 其中 $k=1, 2, \cdots, n-1$ 。

3 应用示例

人均发电量是衡量一个国家发展水平和人民生活水平的一个重要指标。下表是根据 1980-1998 年我国人均发电量的 GM(1,1) 模型, 并预测了 1999-2001 年的人均发电量。

年份	序号 k	原始数据 $x^{(0)}(k)$	梯形公式得到的 $\hat{x}^{(0)}$	相对误差
1980	1	306.35	306.35	0%
1981	2	311.2	301.192	-3.21%
1982	3	324.9	323.041	-0.57%
1983	4	343.4	346.476	0.90%
1984	5	361.61	371.61	2.77%
1985	6	390.76	398.568	2.00%
1986	7	421.36	427.482	1.45%
1987	8	458.75	458.492	-0.06%
1988	9	494.9	491.753	-0.64%
1989	10	522.78	527.426	0.89%
1990	11	547.22	565.687	3.37%
1991	12	588.7	606.724	3.06%
1992	13	647.18	650.738	0.55%
1993	14	712.34	697.944	-2.02%
1994	15	778.32	748.576	-3.82%
1995	16	835.31	802.88	-3.88%
1996	17	888.1	861.123	-3.04%
1997	18	923.16	923.592	0.05%
1998	19	939.48	990.592	5.44%
1999	20	988.60	1062.45	7.47%
2000	21	1073.62	1139.53	6.14%
2001	22	1164.29	1222.19	4.97%

上表使用 C++ 运行, 计算了梯形公式得到的预测数据解, 并计算了相对误差。基于 Cotes 公式的预测数据由于高精度等问题受到一些限制。将所建模型进行误差计算, 可见误差波动小, 预测估计效果较为良好, 可以用于预测我国的人均发电量。

参考文献

[1] 李俊峰, 戴文战. 基于插值和 Newton-Cotes 公式的 GM(1,1) 模型的背景值构造新方法与应用 [J]. 系统工程理论与实践, 2004(10):122-126.