第8卷　第1期　　光学　精密工程　　Vol. 8, No. 1
2000年2月　　OPTICS AND PRECISION ENGINEERING　　Feb., 2000

文章编号　1004-924X(2000)01-0087-04

累积法的基本原理及其在测量数据处理中的应用

石照耀[1], 谢华锟[1], 费业泰[2]
(1. 成都工具研究所, 四川 成都　610056; 2. 合肥工业大学测控系, 安徽 合肥　230009)

摘要: 基于累积算子及其算法, 论述了累积法的基本原理; 以传感器误差修正为例, 介绍了该方法在测量数据处理中的应用。这种新颖的数据处理方法具有简单、直观、不直接处理模型的误差项等特点, 特别适合于多元、高次或多项式模型的参数估计以及时序数据分析。

关　键　词: 累积法; 数据处理; 参数估计; 最小二乘法
中图分类号: O241.5　　**文献标识码**: A

最小二乘法在工程技术中有着广泛深入的应用, 也是精密测量领域中最基本的数据处理方法。最小二乘法以处理模型的剩余误差为出发点, 以剩余误差的平方和最小为目标。本文介绍的累积法是一种新颖的数据处理方法, 它可以保证平均剩余误差为零, 并同时确保绝对误差最小。其特点是简单、直观, 不直接处理模型的误差项。可以说, 在测试技术领域, 这是一种具有较好应用前景的新方法。

1　累积和的概念及算法

1.1　累积和及其算子

所谓累积和就是对原始数据序列按一定的叠加规律, 进行相应的叠加后所得的结果。对于数据序列 $\{x_i: i = 1, 2, \cdots, n\}$, 其各阶累积和 $\sum_{i=1}^{n}{}^{(1)}x_i$, $\sum_{i=1}^{n}{}^{(2)}x_i$, …的定义如下:

$$\sum_{i=1}^{n}{}^{(1)}x_i = x_1 + x_2 + \cdots + x_n = \sum_{i=1}^{n}x_i$$

$$\sum_{i=1}^{n}{}^{(2)}x_i = x_1 + (x_1 + x_2) + (x_1 + x_2 + x_3) + \cdots + (x_1 + x_2 + \cdots + x_n) = \sum_{i=1}^{n}\sum_{j=1}^{i}{}^{(1)}x_j$$

$$\sum_{i=1}^{n}{}^{(3)}x_i = x_1 + [x_1 + (x_1 + x_2)] + [x_1 + (x_1 + x_2) + (x_1 + x_2 + x_3)] + \cdots + [x_1 + (x_1 + x_2) + (x_1 + x_2 + x_3) + \cdots + (x_1 + x_2 + \cdots + x_n)] = \sum_{i=1}^{n}\sum_{j=1}^{i}{}^{(2)}x_j$$

依此类推, 对于任意自然数 k, 可以定义

$$\sum_{i=1}^{n}{}^{(k)}x_i = \sum_{i=1}^{n}\sum_{j=1}^{i}{}^{(k-1)}x_j \tag{1}$$

此式即为求取各阶累积和的计算通式, 其中 $\sum_{i=1}^{n}{}^{(1)}x_i$ 叫做一阶累积算子, $\sum_{i=1}^{n}{}^{(2)}x_i$ 叫做二阶累积算子, …, $\sum_{i=1}^{n}{}^{(k)}x_i$ 也就叫做 k 阶累积算子。当 $k \geq 2$, 称 $\sum_{i=1}^{n}{}^{(k)}x_i$ 为高阶累积算子。

1.2　算法

分析(1)式, 我们便会发现, 按(1)式计算 k 阶算子, 随着数据个数 n 和阶数 k 的增加, 其计算量将呈海量增加。一般情况下, 这种烦琐的计算是不便于实际应用的。因此, 寻找一种简便的计算方法是促进"累积法"走向实用的关键。

下式(2)给出了一种高效计算 k 阶算子的新算法[1]。

$$\sum_{i=1}^{n}{}^{(k)}x_i = \frac{1}{(k-1)!}\sum_{i=1}^{n}(n-i+1)(n-i+2)\cdots(n-i+k-1)x_i \tag{2}$$

1.3　基本累积和

当原始数据的元素全为1时, 称 $\sum_{i=1}^{n}{}^{(k)}1$ 为 k 阶基本累积和, 简记为 $\sum_{i=1}^{n}{}^{(k)}$。对于任意自然数 k, 由(2) 式可求出:

收稿日期: 1999-05-26; **修订日期**: 1999-08-23

$$\sum_{i=1}^{n}{}^{(k)} = \frac{1}{k!}n(n+1)\cdots(n+k-1) \qquad (3)$$

例如: $\sum_{i=1}^{16}{}^{(1)} = 16, \sum_{i=1}^{16}{}^{(2)} = 136, \sum_{i=1}^{16}{}^{(9)} = 1307504$

2　累积法的基本原理

这里以多元线性回归模型的参数估计为例, 介绍累积法的基本原理。至于非线性或者多项式模型的情形, 其基本原理相同。

多元线性回归考虑的是因变量 y 与多个因素(自变量) x_1、x_2, …, x_m 之间的线性关系:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m + \epsilon \qquad (4)$$

其中 $\beta_0, \beta_1, \beta_2, \cdots \beta_m$ 是 $m+1$ 个待估参数; y 是随机变量, 而 x_1、x_2、$\cdots x_m$ 可以是随机变量, 也可以是非随机变量; ϵ 是随机误差。

为了估计模型参数 $\beta_0, \beta_1, \beta_2, \cdots \beta_m$, 对变量进行了 n 次观察, 得到 n 组观察数据 $\{(x_{i_1}、x_{i_2}、\cdots, x_{im}; y_i)\, i = 1, 2, \cdots, n\}$, 一般要求 $n > m$。

由于模型参数个数为 $m+1$ 个, 为求出这些参数, 则累积算子的最高阶数 k 不应小于 $m+1$, 通常取 $k = m+1$。对(4) 实施累加算子, 得到下面的累积法方程组(5):

$$\left.\begin{aligned}
\sum_{i=1}^{n}{}^{(1)} y_i &= \beta_0 \sum_{i=1}^{n}{}^{(1)} + \beta_1 \sum_{i=1}^{n}{}^{(1)} x_{i1} + \cdots \\
&\quad + \beta_m \sum_{i=1}^{n}{}^{(1)} x_{im} + \sum_{i=1}^{n}{}^{(1)} \epsilon_i \\
\sum_{i=1}^{n}{}^{(2)} y_i &= \beta_0 \sum_{i=1}^{n}{}^{(2)} + \beta_1 \sum_{i=1}^{n}{}^{(2)} x_{i1} + \cdots \\
&\quad + \beta_m \sum_{i=1}^{n}{}^{(2)} x_{im} + \sum_{i=1}^{n}{}^{(2)} \epsilon_i \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
\sum_{i=1}^{n}{}^{(m+1)} y_i &= \beta_0 \sum_{i=1}^{n}{}^{(m+1)} + \beta_1 \sum_{i=1}^{n}{}^{(m+1)} x_{i1} + \cdots \\
&\quad + \beta_m \sum_{i=1}^{n}{}^{(m+1)} x_{im} + \sum_{i=1}^{n}{}^{(m+1)} \epsilon_i
\end{aligned}\right\} \qquad (5)$$

记

$$Y = \begin{bmatrix} \sum_{i=1}^{n}{}^{(1)} y_i \\ \sum_{i=1}^{n}{}^{(2)} y_i \\ \vdots \\ \sum_{i=1}^{n}{}^{(m+1)} y_i \end{bmatrix}$$

$$X = \begin{bmatrix} \sum_{i=1}^{n}{}^{(1)} & \sum_{i=1}^{n}{}^{(1)} x_{i1} & \cdots & \sum_{i=1}^{n}{}^{(1)} x_{im} \\ \sum_{i=1}^{n}{}^{(2)} & \sum_{i=1}^{n}{}^{(2)} x_{i1} & \cdots & \sum_{i=1}^{n}{}^{(2)} x_{im} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^{n}{}^{(m+1)} & \sum_{i=1}^{n}{}^{(m+1)} x_{i1} & \cdots & \sum_{i=1}^{n}{}^{(m+1)} x_{im} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix} \qquad \epsilon = \begin{bmatrix} \sum_{i=1}^{n}{}^{(1)} \epsilon_i \\ \sum_{i=1}^{n}{}^{(2)} \epsilon_i \\ \vdots \\ \sum_{i=1}^{n}{}^{(m+1)} \epsilon_i \end{bmatrix}$$

那么, (5) 式简写为:

$$Y = X\beta + \epsilon \qquad (6)$$

上式中 X 是已知的 $(m+1)$ 阶常数方阵, β 是未知的参数向量, 而 ϵ 是均值为零的随机变量, 满足条件

$$E[\epsilon] = 0 \quad \mathrm{cov}(\epsilon, \epsilon) = \sigma^2 I_{m+1} \qquad (7)$$

由(5) 和(6) 式可得到:

$$E[Y] = X\beta \quad cov(Y, Y) = \sigma^2 I_{m+1} \qquad (8)$$

如果 X 是非奇异矩阵, 且 X^{-1} 存在, 那么正规方程为

$$Y = X\beta \qquad (9)$$

由上式得 β 的累积法估计为

$$\hat{\beta} = X^{-1}Y \qquad (10)$$

由累积法估计的 $\hat{\beta}$ 具有这性质: $\hat{\beta}$ 是 β 的无偏估计; $\hat{\beta}$ 是 X 的线性函数, 即 $\hat{\beta}$ 是一线性估计。

由(10) 得到回归方程:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_m x_m \qquad (11)$$

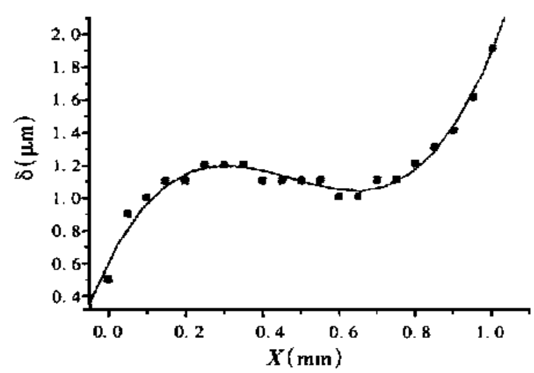

Fig. 1　Fitting curve based on cumulative-sum method

3　应用实例

为修正某光栅式线性测头传感器的系统误差, 对其进行了定标测量。该测头的量程为 1mm, 每隔 0.05mm 测量一点, 其结果如表 1 所列。现用 3 次多项式去拟合系统误差。误差 δ与行程 x 的关系设为:

$$\delta = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \epsilon \qquad (12)$$

Table 1　Results of measurement of a transducer

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
$X_{(i)}$ mm	0	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9	0.95	1.0
$\delta_{(i)}$ μm	0.5	0.9	1.0	1.1	1.1	1.2	1.2	1.2	1.1	1.1	1.1	1.1	1.0	1.0	1.1	1.1	1.2	1.3	1.4	1.6	1.9

X —Displacement; δ—error

应用累积法估计上式的参数值。根据前面给出的公式, 在计算机上利用Matlab 编程, 能十分方便地求出结果

其中

$$\sum_{i=1}^{21}{}^{(1)}\delta = 24.2,\quad \sum_{i=1}^{21}{}^{(2)}\delta = 242.4,\quad \sum_{i=1}^{21}{}^{(3)}\delta = 1784.1,\quad \sum_{i=1}^{21}{}^{(4)}\delta = 10375.4$$

$$\sum_{i=1}^{21}{}^{(1)}x = 10.5,\quad \sum_{i=1}^{21}{}^{(2)}x = 77,\quad \sum_{i=1}^{21}{}^{(3)}x = 442.75,\quad \sum_{i=1}^{21}{}^{(4)}x = 2125.2$$

$$\sum_{i=1}^{21}{}^{(1)}x^2 = 7.175,\quad \sum_{i=1}^{21}{}^{(2)}x^2 = 40.425,\quad \sum_{i=1}^{21}{}^{(3)}x^2 = 190.383,\quad \sum_{i=1}^{21}{}^{(4)}x^2 = 779.24$$

$$\sum_{i=1}^{21}{}^{(1)}x^3 = 5.512,\quad \sum_{i=1}^{21}{}^{(2)}x^3 = 25.43,\quad \sum_{i=1}^{21}{}^{(3)}x^3 = 102.054,\quad \sum_{i=1}^{21}{}^{(4)}x^4 = 365.837$$

而 $\sum_{i=1}^{21}{}^{(1)} = 21$, $\sum_{i=1}^{21}{}^{(2)} = 231$, $\sum_{i=1}^{21}{}^{(3)} = 1771$, $\sum_{i=1}^{21}{}^{(4)} = 10626$

由此可写出矩阵:

$$\mathrm{X} = \begin{bmatrix} 21 & 10.5 & 7.175 & 5.512 \\ 231 & 77 & 40.425 & 25.43 \\ 1771 & 442.75 & 190.383 & 102.054 \\ 10626 & 2125.2 & 779.24 & 365.837 \end{bmatrix}$$

$$\delta = \begin{bmatrix} 24.2 \\ 242.4 \\ 1784.1 \\ 10375.4 \end{bmatrix}$$

由(10)式可求出

$$\hat{\beta} = \mathrm{X}^{-1}\delta$$

$$= [0.6018,\ 4.6320,\ -11.2263,\ 7.8869]^T$$

用累积法拟合出的曲线如图 1 所示, 其标准差 $\sigma = 0.0448$, 拟合方程为

$$\hat{\delta} = 0.6018 + 4.6320x - 11.2263x^2 + 7.8869x^3 \qquad (13)$$

对于(12)式, 按最小二乘法求解, 得到的拟合方程为(14)式, 其标准差 $\sigma = 0.0486$。

$$\hat{\delta} = 0.6018 + 4.6319x - 11.2261x^2 + 7.8867x^3 \qquad (14)$$

比较(13)、(14)两式, 可以看出: 在本例中, 累积法与最小二乘法的求解效果是一致的。

4　结　束　语

由于累积法具有简单、直观、不直接处理模型的误差项、又便于计算机实现等特点, 因而在测量数据处理领域有较好的应用前景。本文介绍的参数估计实例仅仅是其应用的一个方面, 还有许多其它方面的应用值得去探索。

参考文献:

[1]　曹定爱等. 累积法引论[M]. 北京: 科学出版社, 1999.

[2]　陈德钊. 多元数据处理[M]. 北京: 化学工业出版社, 1998.

Fundamental of cumulative-sum method and its application in measurement data processing

SHI Zhao-yao[1], XIE Hua-kun[1], FEI Ye-tai[2]

(1. *Chengdu Tool Research Institute, Chengdu* 610056, *China*; 2. *Dept. of Measurement and Control, Hefei University of Technology, Hefei* 230009, *China*)

Abstract: Based on the cumulative operator and its algorithm, fundamental principle of Cumulative-sum Method is presented. Taking error modification of a transducer for instance, the application of the method is introduced in the field of measurement data processing. This novel data-processing method features simplicity, perception, and non-direct treatment of residual error of models, and is especially employed to estimate parameters of multi-variable high-order non-linear models or to analyze sequential data.

Key words: cumulative-sum method; data processing; parameter estimation; least-square method

作者简介: 石照耀(1964-), 男, 湖南岳阳人。研究员, 博士生, 中国机械工程学会、计量测试学会高级会员, 全国齿轮标准委员会会员。84 年合肥工业大学精仪专业毕业, 获学士学位。88 年西安理工大学量仪专业毕业, 获硕士学位; 89～90 年在英国 Cranfield 理工学院做研究工作。长期从事科研与管理工作, 参与了一系列大型精密仪器的研制; 作为总设计者, 主持研制了我国第一台 CNC 大型滚刀测量机。目前的研究方向为①精密仪器的现代精度理论; ②卧式圆柱坐标测量机。迄今, 参与 3 本论著写作, 发表论文 33 篇。