

观察数据用联系数表示的最小二乘法及应用

郑建青

(宁波大学 理学院, 浙江 宁波 315211)

摘要: 通常观察数据由观察值和观察误差两部分组成. 因此, 可以用联系数作出客观描述, 在此基础上给出了一种新的最小二乘法, 比传统的最小二乘法能更细致地分析偏差的平方和, 从而更精确地得到经验公式.

关键词: 最小二乘法; 观察数据; 偏差; 联系数

中图分类号: O212.1

文献标志码: A

文章编号: 1001-5132 (2013) 01-0057-03

最小二乘法是一种实用的数学优化方法, 通过偏差的平方和最小化, 找出相关数据之间的联系, 建立最佳匹配的函数关系. 通过这些函数关系可以方便地预测出各参数间的最优数据值. 最小二乘法还可以用于曲线拟合等^[1]. 因此, 最小二乘法在科学研究、工农业生产管理等现实生活的各个领域都有应用.

传统的最小二乘法中用到的数据是在实数域内的确定值, 文献[2]把传统的最小二乘法推广到复数域内, 形成所谓的复最小二乘法, 但是这2种方法用到的数据都是以确定值表示的. 由于测量工具的精度和观察方法的限制, 实际科学研究与生产实践中得到的观察数据大都是带有一定程度不确定性的数据, 即有一定的误差. 笔者利用集对分析中的联系数^[3]来表示这种观察数据, 得到了观察数据用联系数表示时的最小二乘法.

1 联系数及其基本运算

本文取以下形式的联系数^[4]: $u = A + Bi$, 其中, $A, B \in R^+$, $i \in [-1, 1]$, A 称为同部, B 称为异部, 由于 i 在区间 $[-1, 1]$ 取值的不确定性, 联系数 $u = A + Bi$ 称为不确定性联系数.

设有2个联系数 $u_1 = A_1 + B_1 i$, $u_2 = A_2 + B_2 i$, 定义 u_1 与 u_2 相等, 它们的同部与异部分别相等, 即 $u_1 = u_2$ 等价于 $A_1 = A_2$, $B_1 = B_2$. u_1 与 u_2 的和是

一个联系数, 记 $u = u_1 + u_2 = A + Bi$, 其中 $A = A_1 + A_2$, $B = B_1 + B_2$, 即联系数的加法同部与异部分别作加法. u_1 与 u_2 的差是一个联系数, 记 $u = u_1 - u_2 = A + Bi$, 其中 $A = A_1 - A_2$, $B = B_1 - B_2$, 即2个联系数作减法, 同部与异部分别作减法运算. u_1 与 u_2 的乘积是一个联系数, 记 $u = u_1 u_2 = A + Bi$, 其中 $A = A_1 A_2$, $B = A_1 B_2 + A_2 B_1 + B_1 B_2$. u_1 与 u_2 的商是一个联系数, $u = u_1 / u_2 = A + Bi$, 其中 $A = A_1 / A_2$, $B = (B_1 - A_1 B_2 / A_2) / (A_2 + B_2)$.

由于二元联系数的值与不确定数 i 的取值有关^[5-7], 因而二元联系数的比较大小也与 i 的取值有关. 若 $i \in I \subseteq [-1, 1]$ 时, 有 $u_1 - u_2 > 0$, 则在 $i \in I \subseteq [-1, 1]$ 时, 有 $u_1 > u_2$; 若 $i \in I \subseteq [-1, 1]$ 时, 有 $u_1 - u_2 < 0$, 则在 $i \in I \subseteq [-1, 1]$ 时, 有 $u_1 < u_2$; 若 $i \in I \subseteq [-1, 1]$ 时, 有 $u_1 - u_2 = 0$, 则在 $i \in I \subseteq [-1, 1]$ 时, 有 $u_1 = u_2$.

2 观察数据用联系数表示的最小二乘法

观察数据用联系数表示的最小二乘法在程序、格式及计算等方面均与传统的最小二乘法基本相同, 只是在有关数学运算中采用了联系数的运算规则. 下面以2个参数的最小二乘法为例.

假设观察数据 $(t_1, y_1), (t_2, y_2), \dots, (t_n, y_n)$ 符合直线 $y = kt + l$, 其中 $t_m = A_{1m} + B_{1m} i$, $y_m = A_{2m} + B_{2m} i$ ($m = 1, 2, \dots, n$) 以联系数表示, 因为所有这些观察数据

(t_m, y_m) 并不完全落在这条直线上, 所以要求选取 k, l , 使得 $y = kt + l$ 在 t_m 处的函数值 $kt_m + l$ 与观察值 y_m 的差都很小, 记偏差 $\delta_m = (kt_m + l) - y_m$, 即要使 δ_m 很小, 可考虑选取常数 k, l , 使 δ_m^2 的和

$$f(k, l) = \sum_{m=1}^n \delta_m^2 = \sum_{m=1}^n [(kt_m + l) - y_m]^2 \text{ 最小.}$$

令

$$\begin{cases} \frac{\partial f}{\partial k} = 2 \sum_{m=1}^n [(kt_m + l) - y_m] t_m = 0, \\ \frac{\partial f}{\partial l} = 2 \sum_{m=1}^n [(kt_m + l) - y_m] = 0, \end{cases}$$

整理合并, 把未知数 k, l 分离出来, 得:

$$\begin{cases} (\sum_{m=1}^n t_m^2)k + (\sum_{m=1}^n t_m)l = \sum_{m=1}^n t_m y_m, \\ (\sum_{m=1}^n t_m)k + nl = \sum_{m=1}^n y_m. \end{cases}$$

设上述线性方程组的解为:

$$k = \frac{n(\sum_{m=1}^n t_m y_m) - (\sum_{m=1}^n t_m)(\sum_{m=1}^n y_m)}{n(\sum_{m=1}^n t_m^2) - (\sum_{m=1}^n t_m)^2},$$

$$l = \frac{(\sum_{m=1}^n t_m^2)(\sum_{m=1}^n y_m) - (\sum_{m=1}^n t_m y_m)(\sum_{m=1}^n t_m)}{n(\sum_{m=1}^n t_m^2) - (\sum_{m=1}^n t_m)^2},$$

与传统的最小二乘法不同, k, l 是同异型联系数, 偏差的平方和为:

$$f(k, l) = \sum_{m=1}^n [(kt_m + l) - y_m]^2,$$

也是同异型联系数, 可以视具体问题取不确定数 $i \in [-1, 1]$ 的值 i_0 , 使得 $f(k, l)$ 最小, 记

$$k|_{i=i_0} = k^*, \quad l|_{i=i_0} = l^*,$$

进而得到经验公式 $y = k^*t + l^*$.

3 应用举例

为了便于与传统的最小二乘法相比较, 取文献[8]的一个例子作适当补充后来说明观察数据用联系数表示的主要结果.

为了测定刀具的磨损速度, 做这样的实验: 经过一定的时间(如每隔 1 h), 测量一次刀具的厚度, 得到一组实验数据(表 1).

表 1 测量结果

顺序	时间/h	刀具厚度/mm	顺序	时间/h	刀具厚度/mm
0	0	27	4	4	26.1
1	1	26.8	5	5	25.7
2	2	26.5	6	6	25.3
3	3	26.3	7	7	24.8

根据生产厂家提供的产品说明书, 时间 t 的不确定度为 0.018, 刀具厚度值不准的变换半范围为示值的 0.4%, 也即平均值的 0.4%, 可得用联系数表示的实验数据(表 2).

表 2 用联系数表示的测量结果

顺序	时间/h	刀具厚度/mm	顺序	时间/h	刀具厚度/mm
0	0.018i	27.0+0.3040i	4	4+0.018i	26.1+0.1044i
1	1+0.018i	26.8+0.1072i	5	5+0.018i	25.7+0.1028i
2	2+0.018i	26.5+0.1060i	6	6+0.018i	25.3+0.1012i
3	3+0.018i	26.3+0.1052i	7	7+0.018i	24.8+0.0992i

解 按照方程组:

$$\begin{cases} (\sum_{m=1}^n t_m^2)k + (\sum_{m=1}^n t_m)l = \sum_{m=1}^n t_m y_m, \\ (\sum_{m=1}^n t_m)k + nl = \sum_{m=1}^n y_m, \end{cases}$$

计算系数, 得具体方程组为:

$$\begin{cases} (140 + 1.0105i)k + (28 + 0.144i)l = 711 + 5.6114i, \\ (28 + 0.144i)k + 8l = 208.5 + 1.03i, \end{cases}$$

解此方程组得:

$$k = -0.3036 - 0.0420i, \\ l = 27.125 + 0.0281i,$$

于是得:

$$f(t) = (-0.3036 - 0.420i)t + (27.125 + 0.0281i),$$

再计算 $t_m (m=0, 1, \dots, 7)$ 处的偏差 $\delta_m = f(t_m) - y_m$ ($m=0, 1, \dots, 7$) 的平方和, 得:

$$f(k, l) = \sum_{m=0}^7 \delta_m^2 = 0.1082 - 0.0008i.$$

笔者用专家估计法, 取 $i=0.8$, 得 $f(k^*, l^*) = 0.1076$, 其中 $k^* = -0.3372$, $l^* = 27.3508$, 于是经验公式为 $y = -0.3372t + 27.3507$. 文献[8]的计算结果为: 偏差的平方和为 0.108165, 经比较本文的

偏差平方和较小, 因而推知经验公式更精确.

4 结论

采用联系数的方法, 可以使最小二乘法用来处理一组带有不确定的实际测量得到的数据, 从这些数据中寻求变量之间的函数关系. 由于在现实生活中的各类测量得到的数据, 大部分存在误差, 即有一定的不确定性, 因此观察数据用联系数 $A + Bi$ 表示. 据此给出基于观察数据用联系数表示的最小二乘法更符合客观实际. 此方法一方面是对传统最小二乘法的创新, 同时也将促进集对分析联系数的理论和应用研究的拓展.

参考文献:

[1] 希洛夫. 最小二乘法[M]. 北京: 测绘出版社, 1955.

- [2] 谷湘潜, 康红文, 曹鸿兴. 复数域内的最小二乘法[J]. 自然科学进展, 2006, 16(1):49-54.
- [3] 赵克勤. 集对分析及其初步应用[M]. 杭州: 浙江科学技术出版社, 2000:1-35.
- [4] 赵克勤. 二元联系数 $A + Bi$ 的理论基础与基本算法及在人工智能中的应用[J]. 智能系统学报, 2008, 6(12): 476-486.
- [5] 唐耀平. 联系数中差异度系数 i 的专家估测法及应用[J]. 数学的实践与认识, 2009, 39(5):67-70.
- [6] 吴亭. 联系数 $a + bi$ 中的不确定数的一个取值公式[J]. 科技通报, 2008, 24(5):595-597.
- [7] 余国祥. 对联系数中的不确定数 i 的研究[J]. 辽宁师范大学学报, 2002, 25(4):349-352.
- [8] 同济大学应用数学系. 高等数学(下册)[M]. 6版. 北京: 高等教育出版社, 2007:124-127.

Least Square Method and its Application Using Connection Number Obtained with Observation Data

ZHENG Jian-qing

(Faculty of Science, Ningbo University, Ningbo 315211, China)

Abstract: The observation data is composed of observation value and observation error, so it can objectively be expressed by connection number. A new least square method is presented on the basis of this stated fact. It can analyze the sum of squares of deviations in a more detailed fashion, and be used to obtain the empirical formula more accurately than the traditional ones.

Key words: least square method; observation data; a deviation; connection number

(责任编辑 史小丽)