西北大学学报(自然科学版)
2006年 6月,第 36卷第 3期, Jun., 2006, Vol. 36, No. 3
Journal of Northwest University (Natural Science Edition)

最小二乘法的创立及其思想方法

贾小勇[1,2], 徐传胜[1], 白 欣[1,3]

(1. 西北大学 数学与科学史研究中心, 陕西 西安 710069; 2. 重庆文理学院 数学与计算机科学系, 重庆 永川 402168;
3. 内蒙古财经学院 公共事业管理学系, 内蒙古 呼和浩特 010051)

摘要: **目的** 探讨最小二乘法的历史发展过程及其创立者的思想与方法。**方法** 历史考察与数理分析。**结果** 勒让德在先驱者解线性方程组的基础上,以整体的思想方法创立了最小二乘法;高斯由寻找随机误差函数为突破,以独特的概率思想导出了正态分布,详尽地阐述了最小二乘法的理论依据。**结论** 两位数学大师异曲同工地谱写了数理统计学的新篇。相比之下,高斯把最小二乘法推进得更远、更深刻,这极大地促进了数理统计学的发展。

关 键 词: 数理统计学; 最小二乘法; 概率; 正态分布; 误差函数
中图分类号: O11 **文献标识码**: A **文章编号**: 1000-274X(2006)03-0507-05

最小二乘法是提供"观测组合"的主要工具之一,它依据对某事件的大量观测而获得"最佳"结果或"最可能"表现形式。如已知两变量为线性关系 $y = a + bx$, 对其进行 $n\ (n > 2)$ 次观测而获得 n 对数据。若将这 n 对数据代入方程求解 a, b 之值则无确定解。最小二乘法提供了一个求解方法,其基本思想就是寻找"最接近"这 n 个观测点的直线。最小二乘法不仅是 19世纪最重要的统计方法,而且还可以称为数理统计学之灵魂。相关回归分析、方差分析和线性模型理论等数理统计学的几大分支都以最小二乘法为理论基础。作为其进一步发展或纠正其不足而采取的对策,不少近现代的数理统计学分支也是在最小二乘法基础上衍生出来的。正如美国统计学家斯蒂格勒(S. M. Stigler)所说,"最小二乘法之于数理统计学犹如微积分之于数学"[1]。

最小二乘法创立的历史过程充满着丰富的科学思想,这些对今日的数学创造仍有着重要的启示意义。本文旨在全面认识最小二乘法的历史系统发育过程以及创立者的思路。

1 先驱者的相关研究

天文学和测地学的发展促进了数理统计学及其他相关科学的发展。丹麦统计史家哈尔德曾指出天文学在数理统计学发展中所起的作用。"天文学自古代至 18世纪是应用数学中最发达的领域。观测和数学天文学给出了建立数学模型及数据拟合的最初例子,在此种意义下,天文学家就是最初的数理统计学家。天文学的问题逐渐引导到算术平均,以及参数模型中的种种估计方法,以最小二乘法为顶峰。"这也说明了最小二乘法的显著地位。

有关统计计算思想记载的著作要首推天文学家罗杰柯茨的遗作,即 1715年其所发论文中所蕴含的统计方法,亦即对各种观测值赋予加权后求其加权平均。尽管当时得到认可,然而事实证明如此计算的结果不太精确[2]。

1749年,欧拉(L. Euler, 1707—1783)在研究木星和土星之间相互吸引力作用对各自轨道影响时,最后得到一个含 8个未知量 75个方程的线性方程组。欧拉的求解方法繁杂而奇特,只能看作是一次尝试。

1750年,天文学家梅耶(T. Meiyer, 1723—1762)通过对月球表面上某定点的观测,得到一含 3个未知数 27个方程的线性方程组。以其中一个方程系数为准,按各方程中此系数的大小分组,较大的 9个、较小的 9个和剩下的 9个分别组成一组。每

收稿日期: 2005-06-20
基金项目: 国家自然科学基金项目(10471111)
作者简介: 贾小勇(1970-),男,甘肃通渭人,西北大学博士生,重庆文理学院讲师,从事数学史研究。

© 1994-2007 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

组内的 9个方程相加,得到一个方程。由得到的 3个方程而求解 3个未知数。梅耶认为,如此所得解之误差比任意选 3个方程而求解之误差要小得多,仅为其 3/27 = 1/9(实际上应为 $1/\sqrt{9} = 1/3$)。由此他得出解类似方程组的一套系统方法,并曾一度相当流行。直到 1760年,罗杰博斯科维奇 (1711—1787)在研究地球真实形状的有关问题时才指出其不足。他认为梅耶确定方程组解的方法还不够精确,应充分满足实际准则,其中包括把一组观测值代入方程组时所产生误差的绝对值之和极小化准则。

1787年,拉普拉斯 (P. S Laplace, 1749—1827)在研究天文学时,得到一个含有 4个未知数 24个方程的线性方程组。其求解方法与梅耶相似,先把 24个方程编号,然后得出 4个方程,以便解出 4个未知数。这 4个方程依次为:

第一个方程,24个方程之和;

第二个方程,前 12个方程之和减去后 12个方程之和;

第三个方程,编号为 3, 4, 10, 11, 17, 18的方程之和减去编号 1, 7, 14, 20的方程之和;

第四个方程,编号为 2, 8, 9, 15, 16, 21, 22的方程之和减去编号为 5, 6, 12, 13, 19的方程之和。

拉普拉斯并没有给出如此组合的原因,但可以看到如此组合可使同一个方程至少被使用两次,而前 22个方程被使用三次。这已比前述结果前进了一大步。由此可见,早期的数学家们致力于组合方程而忽视了整体的均衡性[3]。纵观数学史,在每一新的理论创立之前,总是离不开先驱者的努力。这些先驱者的见解往往不完美,甚至含有漏洞和缺陷,但他们的工作对于新思想、新理论的创立是十分必要的。

2 勒让德创立最小二乘法

勒让德 (A. M. Legendre, 1752—1833)是法国军事学校的教授,曾任多届政府委员,后来成了多科工艺学校的总监,直至 1833年逝世。他一直保持热情而有规律的数学研究工作,由于解决了许多类型的的问题,其名字常存于许多定理之中。数学史家克莱因 (M. Kline, 1908—1992)认为勒让德之所以名列拉格朗日 (J. L. Lagrange, 1736—1813)、拉普拉斯、蒙日 (G Monge, 1746—1818)之后,是因为其工作不如这三人深刻。尽管勒让德的工作引起许多重要理论的产生,但这只是在他的研究成果被更强有力的思想接受后才实现的。最小二乘法就是一个典型实例。

最小二乘法最早出现在勒让德 1805年发表的论著《计算彗星轨道的新方法》附录中。该附录占据了这本 80页小册子的最后 9页,在前面关于卫星轨道计算的讨论中没有涉及最小二乘法,可以推测他当时感到这一方法尚不成熟。

勒让德在该书 72～75页描述了最小二乘法的思想、具体做法及其优点。以引进这种方法的理由为开端:“所研究的大多数问题都是由观测值来确定其结果,但这几乎总产生形如 $E = a + bx + cy + fz + \cdots$方程的方程组,其中 $a, b, c, f, \cdots$是已知系数,它们从一个方程到另一个方程是有变动的。$x, y, z, \cdots$是未知的,它们必须根据将每个方程 E化为 0或很小的量来确定”[4]。用现代术语可描述为,一 n个未知量 m个方程的线性方程组 $(m > n)$,

$$E_j = a_{j0} + a_{j1}x_1 + a_{j2}x_2 + \cdots + a_{jn}x_n \quad (j = 1, 2, 3, \cdots m)。$$

寻找“最佳 ”近似解,以使所有 E_j都变小。勒让德认为:“赋予误差的平方和为极小,则意味着在这些误差间建立了一种均衡性,它阻止了极端情形所施加的过分影响。这非常好地适用于揭示最接近真实情形的系统状态 ”[5]。

为了确定误差平方的最小值,勒让德运用了微积分工具。即为使平方和 $\sum_{i=1}^{m} E_i^2 = E_1^2 + E_2^2 + \cdots E_m^2$在 x_i变动时有最小值,则它对 x_i的偏导数必为 0。由此得如下线性方程组

$$\sum_{j=1}^{m} a_{ji}a_{j0} + x_1\sum_{j=1}^{m} a_{ji}a_{j1} + \cdots + x_i\sum_{j=1}^{m} a_{ji}^2 + \cdots + x_n\sum_{j=1}^{m} a_{ji}a_{jn} = 0\ (i = 1, 2, \cdots n)。$$

这样,就得到一含有 n个未知量 n个方程的线性方程组,用“现成的方法 ”是可以解出的。

关于最小二乘法的优点,勒让德指出以下几条:

1) 通常的算术平均值是其特例。即 $n = 1, a_{j1} = -1$时,令 $b_j = a_{j0}$,则误差的平方和为 $(b_1 - x)^2 + (b_2 - x)^2 + \cdots + (b_m - x)^2$。

对其求关于 x的偏导数,则使此和极小的方程是 $(b_1 - x) + (b_2 - x) + \cdots (b_m - x) = 0$,故解为

$$x = \frac{b_1 + b_2 + \cdots + b_m}{m}。$$

它正是 m个观测值的算术平均值。

2)如果观测值全部严格符合某一方程组的要求,则此解必是最小二乘法的解。

3)如果舍弃或增加观测值,则修改所得方程组

© 1994-2007 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

即可。

勒让德的成功在于他从一个新的角度来看待这个问题,不像其前辈那样致力于找出几个方程(个数等于未知数的个数)再去求解,而是考虑误差在整体上的平衡。从某种意义讲,最小二乘法是一个处理观测值的纯粹代数方法。要将其应用于统计推断问题就需要考虑观测值的误差,确定误差分布的函数形式。

3　随机误差的早期研究

天文学家伽利略(G. Galileo, 1564—1642)可能是第一个提出随机误差概念并对其有所研究的学者。他在 1632年出版的著作《关于两个主要世界系统的对话》中提及这个问题。尽管他用"观测误差"这个名称,但所描述的性质实则为现在的随机误差分布。伽利略认为:①所有观测值都可能有误差,其源于观测者、仪器工具及观测条件等;②观测误差对称的分布在 0的两侧,因仪器工具使得观测值比真值大或小的可能性是等同的;③小误差出现的频率大于大误差。由此可见,伽利略所设想的误差分布函数 $f(x)$应满足关于 y轴对称,且随 $|x|$增加而递减等条件。这个定性式讨论的范围,成为日后学者研究这一问题的出发点。

辛普森(Thomas Simpson, 1710—1761)在 1755年向皇家学会宣读的文章《在应用天文学中取若干观测平均值的好处》中试图证明,若以观测值的平均值估计真值,误差将比单个观测值要小,且随着观测次数的增加而减小。辛普森对一种极特殊的误差分布证明了其结论。他假定在一次天文测量中以秒来度量的误差只能取 0, ±1, ±2, ±3, ±4, ±5,这 11个值,取这些值的概率则以在 0处最大,然后在两边按比例下降,直到 ±6处为 0,即

$$p\{X = i\} = (6 - |i|)r \quad i = 0, \pm 1, \pm 2, \cdots \pm 5$$

其中 $r = 1/36$。

根据所给的分布,可算得单个误差不超过 1秒的概率为 $16/36 = 0.444$,不超过 2秒的是 $24/36 = 0.667$。为比较起见,他又计算出 6个误差的平均值不超过 1秒的概率是 0.725,不超过 2秒的是 0.967。易见平均值的估计优于单个值。此可视为第一次从概率角度严格证明算术平均值的优良性。由此出发,辛普森得出了现今熟知的独立均匀分布和的密度函数公式。

拉普拉斯与辛普森的研究途径不同,他直接考虑误差理论的基本问题,即应取怎样的分布为误差分布,以及在确定误差分布后,如何根据未知量 θ的多次测量结果 $\theta_1, \theta_2, \cdots \theta_n$去估计 θ。拉普拉斯可能知道伽利略的有关结论,其给出误差分布 $f(x)$应满足类似的条件:①$f(x) = f(-x)$;②$x \to \infty, f(x) \to 0$(因无限大误差的概率为 0);③$\int_{-\infty}^{\infty} f(x)\,dx = 1$(因在任意两数值之间曲线下方的面积代表观测具有的误差在这两个值之间的概率)。显然,有很多函数满足这三条性质,为确定其一,拉普拉斯作了如下推理:由条件 ②知,随着 x的增加曲线 $f(x)$愈来愈平缓,因而其下降率 $-f'(x)$也应随 x增加而下降。设 $-f'(x) = mf(x), x \geq 0\ m > 0$且为常数,

则可解得 $f(x) = ce^{-mx}, c > 0$且为常数,由 $f(x) = f(-x)$得

$f(x) = ce^{mx}, x < 0$。再由 $\int_{-\infty}^{\infty} f(x)\,dx = 1$,得 $c = m/2$,于是 $f(x) = \frac{m}{2} e^{-m|x|}, -\infty < x < \infty$。

这就是今日的拉普拉斯分布[6]。然而,拉普拉斯很快发现,基于这个误差函数的计算是相当繁杂的,故也就不可能有多大的实际应用价值。后来,拉普拉斯得到一个更加复杂的函数表达式,也只好无功而返了。至此,拉普拉斯已感到无能为力了。今天看来,我们为拉普拉斯深感惋惜,因早在 1780年他就推广了棣莫弗(A. De Moivre, 1667—1754)的结果(棣莫弗在 1730年由二项分布的近似公式导出正态分布密度函数表达式),得到了中心极限定理的较一般形式,但他未将这一成果应用到确定误差分布的问题上来[7]。因此,科学的重大发现往往与人擦肩而过,即使是功力深厚、思维敏捷的大数学家,也是如此。

4　高斯和最小二乘法

德国慕尼黑博物馆的高斯(C. F. Gauss, 1777—1855)油画像下写有:"他的思想深入数字、空间、自然的最深秘密;他测量星体的路径及地球的形状和自然力;他推动了数学的进展直到下个世纪"。的确,高斯是"能以九霄云外的高度按照某种观点掌握星空和深奥数学的天才。"由正态分布的导出可对高斯创造性思维略见一斑。

1809年,高斯发表论著《天体运动理论》。在该书末尾,他写了一节有关"数据结合"的问题,以极其简单的手法导出误差分布——正态分布,并用最小二乘法加以验证。关于最小二乘法,高斯宣称自 1795年以来他一直使用这个原理。这立刻引起了

© 1994-2007 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

勒让德的强烈反击,他提醒说科学发现的优先权只能以出版物确定,并严斥高斯剽窃了他人的发明[8]。他们间的争执延续了多年。因而,这两位数学家之间关于优先权的争论,在数学史上的知名度仅次于牛顿和莱布尼兹之间关于微积分发明权的争论。现在一般认为,二人各自独立地发明了最小二乘法,尽管早在 10年前,高斯就使用这个原理,但第一个用文字形式发表的是勒让德。

高斯较之于勒让德把最小二乘法推进得更远,他由误差函数推导出这个方法并详尽阐述了最小二乘法的理论依据。其推导过程如下[9]:

设误差密度函数为 $f(x)$,真值为 x, n个独立测定值为 $x_1, x_2, \cdots, x_n$,由于观测是相互独立的,因而这些误差出现的概率为

$$L(x) = L(x; x_1, x_2 \cdots x_n) = f(x_1 - x) f(x_2 - x) \cdots f(x_n - x) \quad (1)$$

要找出最有希望的误差函数应使 $L(x)$达极大,高斯认为 $\bar{x}$就是 x的估计值,并使 $L(x)$取得极大值。对式(1)两端取对数得

$$\ln L(x) = \sum_{i=1}^{n} \ln f(x_i - x) \text{。} \quad (2)$$

再对式(2)求导 $\frac{\mathrm{d}\ln L(x)}{\mathrm{d}x} = \sum_{i=1}^{n} \frac{f'(x_i - x)}{f(x_i - x)}$,

记 $g(x) = f'(x)/f(x)$ 则有 $\sum_{i=1}^{n} g(x_i - \bar{x}) = 0$上式求对 x_i的偏导数 $\frac{\partial g}{\partial x_i} + \frac{\partial g}{\partial x_n}\frac{\partial x_n}{\partial x_i} = 0$,

而 $\sum_{i=1}^{n} x_i - n\bar{X} = 0$有 $\frac{\partial x_n}{\partial x_i} = -1 \ (i \neq n)$,

则对任意 i有$\frac{\partial g}{\partial x_i} = \frac{\partial g}{\partial x_n}$,即$\frac{\partial g}{\partial x_i} = c$($c$为常数),可得 $g(x) = cx + b$,

$$\sum_{i=1}^{n} g(x_i - \bar{x}) = \sum_{i=1}^{n} [c(x_i - \bar{x}) + b] = c\sum_{i=1}^{n} (x_i - \bar{x}) + nb = 0\text{。}$$

因 $\sum_{i=1}^{n} (x_i - \bar{x}) = 0$可推得 $b = 0$,则有

$$g(x) = f'(x)/f(x) = cx,$$

积分可得 $f(x) = k\mathrm{e}^{\frac{1}{2}cx^2}$由 $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$, 应有 $c < 0$,取 $c = -\frac{1}{\sigma^2}$, 可得 $k = \frac{1}{\sqrt{2\pi}\sigma}$,则有 $f(x) = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{x^2}{2\sigma^2}}$此即正态分布 $N(0, \sigma^2)$。

这样可知 $(x_1, x_2, \cdots, x_n)$ 的误差密度函数为 $(\sqrt{2\pi}\sigma)^{-n}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - x)^2\right\}$。

要使此式达到极大值,必须选取 $x_1, x_2, \cdots, x_n$之值而使表达式 $\sum_{i=1}^{n}(x_i - x)^2$ 达极小值。于是,可得 $x_1, x_2, \cdots, x_n$的最小二乘法估计。

在推证过程中,高斯有两个创新之处:①他不像其前辈那样,采取贝叶斯式的推理方式,而是直接构造观测值的似然函数,即导出误差函数使其达极大估计量;②高斯用逆向思维来思考这个问题,即先承认算术平均值 $\bar{x}$是所求的估计,即"如果在相同的环境和相等的管理下对任一个量经由多次直接观测确定,则这些观测的算术平均值是最希望要的值"。这是高斯大胆采用了人们千百年来的实际经验,实为高斯之独创性思维。这也正如他所说:"数学,要有灵感,必须接触现实世界"。

最小二乘法在 19世纪初发明后,很快得到欧洲一些国家的天文学家和测地学家的广泛关注。据不完全统计,自 1805年至 1864年的 60年间,有关最小二乘法的研究论文达 256篇,一些百科全书包括 1837年出版的大不列颠百科全书第 7版,亦收入有关方法的介绍。同时,误差的分布是"正态"的,也立刻得到天文学家的关注及大量经验的支持。如贝塞尔(F. W. Bessel, 1784—1846)对几百颗星球作了三组观测,并比较了按照正态规律在给定范围内的理论误差值和实际值,对比表明它们非常接近一致[10]。拉普拉斯在 1810年也给出了正态规律的一个新的理论推导并写入其《分析概论》中。正态分布作为一种统计模型,在 19世纪极为流行,一些学者甚至把 19世纪的数理统计学称为正态分布的统治时代。在其影响下,最小二乘法也脱出测量数据意义之外而发展成为一个包罗极大,应用及其广泛的统计模型。到 20世纪正态小样本理论充分发展后,高斯研究成果的影响更加显著。

综上可知,勒让德和高斯发现最小二乘法是从不同的角度入手的:一个是为解线性方程组,一个是寻找误差函数;一个用的是整体思维,考虑方程组的均衡性,一个用的是逆向思维,首先接受经验事实;一个是纯代数方法,一个致力于应用。相比而言,高斯不愧为数学王子,他把最小二乘法推进得更远、更深刻,这极大地推进了数理统计学的发展。

参考文献:

[1] LANCASTER H O. Encyclopedia of statistical science [J]. New York: Wiley, 1988.

© 1994-2007 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

[2] HACKING I. The Taming of Chance[M]. Cambridge: Cambridge University Press, 1990.
[3] COHEN I B. Revolution in Science[M]. Cambridge: The Belknap Press of Harvard University Press, 1985.
[4] PLACKETT R L. The Discovery of the method of Least Squares[J]. Biometrika, 1972, (59): 239-251.
[5] STIGLER S M. The History of Statistics[M]. Cambridge: Harvard University Press, 1986: 94-96.
[6] TODHUNTER I. A History of the Mathematical of theory of probability from the times of pascal to that of Laplace[M]. New York: Chelsea, 1965: 482-517.
[7] 徐传胜,吕建荣. 亚伯拉罕·棣莫弗的概率思想与正态概率曲线[J]. 西北大学学报:自然科学版, 2006, 36(2): 339-343.
[8] SHEYNIN O B. GAUSS C F and the theory of Error[J]. Archive for History of Exact Science, 1979, (20): 21-72.
[9] WATERHOUSE W C. Gauss's first argument for least squares[J]. Archive for history of Exact Science 1991, (41): 41-52.
[10] VICTOR J K. 数学史通论[M]. 李文林,译. 北京:高等教育出版社, 2004: 586-588.

(编　辑　姚　远)

The invention and way of thinking on least squares

JIA Xiao-yong[1,2], XU Chuan-sheng[1], BAI Xin[1,3]

(1. Research Center for the History of Mathematics and Science, Northwest University, Xi'an 710069, China; 2. Mathematics and Computer Science Department, Chongqing University of Arts and Sciences, Yongchuan, 402168, China; 3. Inner Mongolian Finance and Economics College, Department of Public Administration, Huhhot 010051, China)

Abstract: Aim To explore the historical development of the method of least squares and its inventors' thought approach. **Methods** Historical investigation and mathematical analysis. **Results** Based on solving system of linear equations, A. M. Legender invented the method of least squares with a whole thought; while C. F. Gauss deduced normal distribution with a particular probability thought by looking for random error function and interpreted in detail the theoretical foundation of the method of least squares. **Conclusion** They invented the important method respectively in different way. In comparison, C. F. Gauss developed it further.

Key words: mathematical statistics; least squares method; probability; normal distribution; error function

© 1994-2007 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net