

# 直线拟合的技术综合方法

朱鹤年，朱美红，肖志刚\*

清华大学物理系

周光召先生在1999年10月说：“为什么本世纪最重要的物理学发现又恰恰在德国的土地上发生？”他归纳为三点：首先德国人非常重视实验和实验数据的分析；第二个特点就是德国有很强的数学传统；第三个是德国有很强的哲学传统。次年《物理》杂志刊登该讲话时的题目是《希望在中国产生诺贝尔奖获得者》。光召之论比钱学森之问见报早6年

重视实验及评估, 数理结合, 冲破哲学陈见

2011年1月SCIENCE主编Alberts B.在短文

“《科学》杂志新设的一种教学奖”中，批评了

以往大学本科的科学教育，谈及亲历体会时说：

“既没有掌握对科学本质的真正理解，也没有

获得关于产生和评估科学证据及其解释的

任何经验，而这两点才是现代意义下‘科学教

育’最核心的要素。许多学校至今仍然一成不变

地保留着这种缺憾。”

（转引自葛惟昆教授的译文）

# 1. 前言

勒让德 - 高斯 创立的最小二乘法(LSM), 使残差平方和  $RSS$  (residual sum of square) 极小, LSM 是统计学发展中的里程碑. <sup>[1]</sup> 直线拟合常用 LSM. LSM 的前提是高斯-马尔科夫 (G-M) 假定: 因变量  $y_i$  具有独立、同分布的随机误差. <sup>[2][3]</sup>

我们已用反证法证明: 测量多点散布数据作直线拟合, 散布数据拟合主要目的是为减小  $y_i$  中随着  $x_i$  不同而具随机性的未定系统误差的影响, <sup>[4]</sup> 因此 G-M 假定常不满足

**附录 1: 反证法**: 如因变量  $y_i$  仅有总体标准差为  $\sigma_y$

的随机误差, 则不必测散布数据, 只要集中在首尾两点  $x_{\min}$ ,  $x_{\max}$  处各测  $n/2$  次, 这样即可使回归的因变量  $y_i$  的标准差  $S_y$  极小.

$$\text{首尾各半: } s_{b1}' \approx \sqrt{2} \left( \frac{\sigma_y}{\sqrt{n/2}} \right) / (x_{\max} - x_{\min}) = \left( \frac{2\sigma_y}{\sqrt{n}(x_{\max} - x_{\min})} \right)$$

由拙作《基础物理实验教程》式(5.27):  $n$  组均布 LSM 拟合结果为

$$\text{LSM} \quad s_{b1} \approx s_{b1} \approx \sqrt{3} \left( 1 - \frac{1}{n+1} \right) \cdot \left( \frac{2\sigma_y}{\sqrt{n}(x_{\max} - x_{\min})} \right)$$

集中测首尾两点, 斜率标准差  $s_{b1}$  约为 LSM 的  $1/\sqrt{3}$  倍.

此外, LSM 令  $RSS = \sum v_i^2 = \sum (y_i - \hat{y}_i)^2$  极小而使稳健性 (抗粗差性) 变差; 直线两端的  $y_i$  的误差  $e_i$  对斜率  $b_1$  影响大, 因为  $|\partial b_1 / \partial y_i| \propto |x_i - \bar{x}|$  .

## 最小一乘法

稳健性好, 它以残差绝对值和 SAR (sum of absolute residuals) 极小为依据,  $SAR = \sum |v_i| = \sum |y_i - \hat{y}_i|$ , 也称最小绝对值(和)法, 缩写成 LAM. 由于 LAM 算法较复杂, 它的出现虽然比 LSM 约早 40 年, 但直到 60 年前进入计算机时代才开始被逐步推广.<sup>[1]</sup>

# 简化拟合法

$$\text{偶数 } n' \leq n \text{ 时用式 } b_1' = \frac{\sum_{i=1}^n y_i - \sum_{i=1}^{n'/2} y_i}{\sum_{i=1}^n x_i - \sum_{i=1}^{n'/2} x_i} \text{ 由 } n \text{ 个数据点}$$

求斜率  $b_1'$ ，是前计算机时代国外早已使用的简化拟合法 (SFM)。<sup>[5]</sup>

当高度异常值点出现在直线线段端点附近时，由  $|\partial b_1 / \partial y_i| \approx \text{const} \propto |x_i - \bar{x}|^0$ ，可看出该法稳健性可能优于 LSM。

# 剔除高度异常值

先剔高度异常值再回归是理想的稳健方法，但回归时的统计离群值判别是至今未能系统解决的问题。<sup>[6]</sup> 部分文献中，先用全部  $n$  组数据求出回归直线方程  $\hat{y}_{i,n} = b_{0,n} + b_{1,n}x_i$ ，用最大残差  $|y_i - \hat{y}_{i,n}|_{\max}$  与其标准差  $s_{y_i,n}$  之比  $G = |y_i - \hat{y}_{i,n}|_{\max} / s_{y_i,n}$  作为的离群值判断用的统计量。<sup>[7]</sup> 用  $G > G_{\text{cr},0.99}$  判据剔除“粗差”后，常出现被剔除值的残差处于剩余  $(n-1)$  组数回归直线残差的“包含总体不少于 99.9%、概率为 99% 的统计允许区间”内（依 ISO 及 GB 标准<sup>[8]</sup>）， $|y_i - \hat{y}_{i,n-1}|_{\max} \leq k_2(0.999, 0.99) * s_{y,n-1}$ ，这说明此类判据往往有逻辑性瑕疵。<sup>[4]</sup>

对于理想模型为截距非零的、自由度不小于5的直线拟合问题,我们提出技术综合方法(technical hybrid method,简称 THM),力图部分综合上述几种方法的特点.

不作粗差剔除,仅在求斜率时适度降低  $k_i = |y_i - \hat{y}_{i,n-1}| / s_{y_i,n-1}$  大的个别数据  $y_i$  的权重,以达如下效果:

- a) 稳健性显著优于 LSM 及 SFM, 也优于 LAM;
- b) 残差绝对值和、标准偏差近似与 LAM 相同;
- c) 计算比 LAM 简单.

## 2. 技术综合法的思路简介

第1步, 综合LSM与SFM的特点, 用基于LSM的加权拟合求斜率, 使 $|\partial b_1 / \partial y_i|$ 近似正比于 $|x_i - \bar{x}|^{0.5}$ .

第2步, 对个别 $k_i = |y_i - \hat{y}_{i,n-1}| / s_{y_i,n-1}$ 大于 $k_2(0.95, 0.95)$ <sup>[8]</sup>的可能异常的值, 降低一些权重, 作第2轮加权拟合求出斜率 $b_1$ 与截距初值 $\tilde{b}_0$ .

第3步, 固定 $b_1$ , 在 $\tilde{b}_0 \pm 3s_{\tilde{b}_0}$ 范围内用最小一乘判据求出截距 $b_0$ 来.

# 3. 用蒙特卡洛法建稳健性判据

新方法要有更普遍的适用性与稳健性，我们用蒙特卡洛法（MCM）给出几种方法稳健性的定量统计比较判据。

a) 先构造  $n$  对误差为零、位于理想直线  $y_{it} = \exp(1) + x_i \pi$  的点  $(x_i, y_{it})$ ，这里设  $x_i$  等间隔， $x_i = (2+n)i, i = 1, 2, \dots, n$ 。  $x_i$  在一定区间内呈均匀分布时也能得出与下文类似的结论。

b) 建立  $m$  组  $n-1$  个均值为 0、总体标准差  $\sigma = 0.01$  的正态（或均匀）分布的随机数组  $e_{ij}$  ( $j = 1, 2, \dots, m, m \geq 1000$ )，作为  $n$  个因变量中  $n-1$  个  $y_i$  的误差， $y_i = y_{it} + e_{ij}$ 。

剩余一个因变量的“粗差”取  $e_{\max} = \pm l\sigma$ ， $l$  取 3-16 的正整数，上限取 16 是因为直线拟合中若单个误差超出  $14\sigma$  时统计离群值误判、漏判的概率极低。

c) 针对总体正态分布或均匀分布， $l = |e_{\max} / \sigma|$  取 3-16 的不同数， $n$  取 6~16 等不同点数，“粗差”点分别取在直线上前半段的不同位置，分别用 LSM、LAM、THM 求出直线方程  $\hat{y}_i = b_0 + b_1 x_i$ ，计算下列两类特征量各  $m$  组以作统计处理。

## 残差类特征量:

标准偏差  $s_y = \sqrt{\sum (y_i - \hat{y}_i)^2 / (n-2)}$ , LSM 判据用

残差绝对值均值  $\overline{|v_i|} = \sum |y_i - \hat{y}_i| / n = SAR/n$ , LAM 判据用  
(与 SAR 相比, SAR/n 与  $s_y$  等统计量有相近的量值)

## 新增误差类特征量:

误差绝对值均值  $\overline{|e_i|} = \sum |y_{it} - \hat{y}_i| / n$ , 稳健性判据用

最大误差绝对值  $|e_i|_M = \max(|y_{it} - \hat{y}_i|)$ . 稳健性判据用

d) 在  $l$ 、 $n$  与“粗差”点位置相同时，用  $m$  组误差数据分别计算残差、误差的特征量值，再得出两类统计量

**残差类统计量：**  $m$  组标准差的平均值  $\overline{(s_y)_j}$ ，

$m$  组残差绝对值均值的平均值  $\overline{|v_i|_j}$ ；

**误差类统计量：**  $m$  组误差绝对值均值的平均值  $\overline{|e_i|_j}$ ，

$m$  组最大误差绝对值的平均值  $\overline{|e_i|_{Mj}}$ 。

f) 用上述误差统计量作比较时，采用比[4]中第 2.5 节教学用判据更严的判据。用 MCM 比较不同方法时，若误差类统计量相对差大于 5%，则认为统计量小的方法稳健性优；若相对差小于 5%则认为两法统计量近似相等，稳健性基本相同。

## 附录 2:

何定 5% ?

标准差或不确定度分量合成的微小分量判据中,

$$S_{\text{negligible}} \leq \left( \frac{1}{4} \sim \frac{1}{6} \right) * S_{\text{max}} .$$

$$\sqrt{1^2 + \left( \frac{1}{3.125} \right)^2} = 1 + 5.0\% , \quad \sqrt{1^2 - \left( \frac{1}{3.203} \right)^2} = 1 - 5.0\% .$$

相差超过 5%，等效于方和根合成时有不可忽略分量。

# 4. 误差正态分布时三种直线拟合方法稳健性的 MCM 比较

对于主要误差源自总体正态,

$l = \lfloor e_{\max} / \sigma \rfloor$  取 3-16 的 14 种不同数,

$n$  取 6~16 等不同点数,

粗差点取直线前半段上的不同位置, 共有 882 种组合.

对每个组合分别用 LSM、LAM、THM 直线拟合, 每个组合计算  $m=1000$  次. 考虑到回归时通常要求自由度  $\nu = n - 2 \geq 5$ , [4] 汇总时只计算  $n \geq 7$  的 495 种组合. 两类统计量的相对比较值的结果见表 1、表 2.

表 1A. 三种拟合法的误差类统计量比较

误差统计量比较		$\left(\overline{ e_{i j} }\right)_{\text{THM}} / \left(\overline{ e_{i j} }\right)_{\text{LAM}} - 1$			$\left(\overline{ e_{i j} }\right)_{\text{THM}} / \left(\overline{ e_{i j} }\right)_{\text{LSM}} - 1$		
n	组合数	均值	>5%占	<-5%占	均值	>5%占	<-5%占
6	42	-8.5%	7.1%	69%	-56.6%	0%	100%
7	56	-8.3%	3.6%	77%	-55.3%	0%	93%
8	56	-10.2%	5.4%	82%	-57.8%	0%	95%
9	70	-9.5%	1.4%	87%	-53.5%	0%	93%
10	70	-12.2%	0%	89%	-55.2%	0%	93%
11	84	-11.3%	0%	92%	-52.5%	0%	93%
12	84	-9.9%	0%	90%	-53.0%	0%	93%
13	98	-9.6%	0%	93%	-49.8%	0%	93%
14	98	-17.8%	0%	100%	-42.6%	9%	88%
15	112	-8.6%	0%	92%	-48.7%	0%	93%
16	112	-13.9%	0%	100%	-48.7%	3.6%	92%
6-16	882	-11.2%	1.0%	90.5%	-51.2%	1.5%	92.6%
均匀分布	120组 n=15	-15.4%	0%	100%	-35.2%	17%	78%

对于误差统计量  $\overline{|e_{i|j}|}$  882 种组合 THM 比 LAM 平均小 11.2%，LAM 明显大的组合数占 91%，明显小的组合数仅占 1%；THM 比 LSM 平均小 51%，LSM 明显大的占 93%，明显小的只占 1.5%。

表 1B. 三种拟合法的误差类统计量比较

误差统计量比较		$\left(\overline{ e_i _{M_j}}\right)_{THM} / \left(\overline{ e_i _{M_j}}\right)_{LAM}^{-1}$			$\left(\overline{ e_i _{M_j}}\right)_{THM} / \left(\overline{ e_i _{M_j}}\right)_{LSM}^{-1}$		
n	组合数	均值	>5%占	<-5%占	均值	>5%占	<-5%占
6	42	-9.4%	9.5%	69%	-56.5%	0.0%	98%
7	56	-10.2%	7.1%	75%	-54.5%	0.0%	93%
8	56	-12.2%	5.4%	82%	-57.8%	0.0%	93%
9	70	-12.1%	2.9%	86%	-53.8%	0.0%	93%
10	70	-14.2%	2.9%	89%	-54.8%	0.0%	93%
11	84	-14.5%	1.2%	92%	-52.4%	0.0%	93%
12	84	-12.4%	0.0%	92%	-53.3%	0.0%	93%
13	98	-12.2%	0.0%	94%	-49.8%	0.0%	93%
14	98	-24.4%	0.0%	100%	-46.3%	7.1%	90%
15	112	-12.1%	0.0%	94%	-49.1%	0.0%	92%
16	112	-19.0%	0.0%	100%	-50.0%	0.0%	92%
6-16	882	-14.5%	1.8%	90.7%	-51.8%	0.8%	92.5%
均匀分布 120 组 n=15		-21.5%	0%	100%	-39.2%	15%	82%

对于误差统计量  $\overline{|e_i|_{M_j}}$ ，882 种组合中 THM 比 LAM 平均小 14.5%，明显小的组合数占 91%，明显大的只占 1.8%；THM 比 LSM 平均小 52%，明显小的组合数占 92%，明显大的只占 0.8%。

误差统计量比较		$\left(\overline{ e_i _{Mj}}\right)_{THM} / \left(\overline{ e_i _{Mj}}\right)_{LAM}^{-1}$			$\left(\overline{ e_i _{Mj}}\right)_{THM} / \left(\overline{ e_i _{Mj}}\right)_{LSM}^{-1}$		
n	组合数	均值	>5%占	<-5%占	均值	>5%占	<-5%占
15	正态分布 112	<u>-12%</u>	0%	<u>94%</u>	-49%	0%	92%
n=15	均匀分布 112 组	<u>-21.8%</u>	0%	<u>100%</u>	-43.4%	9%	88%

表 1B 最后一行列出了  $n=15$  时 14 点误差呈  $\sigma=0.01$  的均匀分布时的误差统计量. 因为非粗差点比正态分布密集, THM 的稳健性仍显著地优于 LSM, 也较显著地优于 LAM, 全部组数的统计量比 LAM 至少小 9%. 总体误差近均匀分布时 LAM 的稳健性一般明显优于 THM.

**结论: THM 稳健性优于 LAM,  
并显著地优于 LSM.**

表 2. 三种拟合法的残差类统计量比较

残差统计量比较		$\left(\overline{ v_i _j}\right)_{\text{THM}} / \left(\overline{ v_i _j}\right)_{\text{LAM}} - 1$			$\left(\overline{ v_i _j}\right)_{\text{THM}} / \left(\overline{ v_i _j}\right)_{\text{LSM}} - 1$		
n	组合数	均值	>5%占	<-5%占	均值	>5%占	<-5%占
7~16	840	1.6%	0.0%	0.0%	-13.7%	0.0%	86%
残差统计量比较		$\left(\overline{(s_y)_j}\right)_{\text{THM}} / \left(\overline{(s_y)_j}\right)_{\text{LAM}} - 1$			$\left(\overline{(s_y)_j}\right)_{\text{THM}} / \left(\overline{(s_y)_j}\right)_{\text{LSM}} - 1$		
n	组合数	均值	>5%占	<-5%占	均值	>5%占	<-5%占
7~16	840	0.31%	0.6%	0%	7.3%	54%	0.0%

从表 2 可看出， $\overline{|v_i|_j}$  是 LAM 极小，LAM 与 THM 两法之差全部在 ±5% 以内，说明其残差特征量近似相同。

$\overline{(s_y)_j}$  是 LSM 极小，LAM 与 THM 两法之差有 835 个组合在 ±5% 以内，全部量的均值 0.31%。

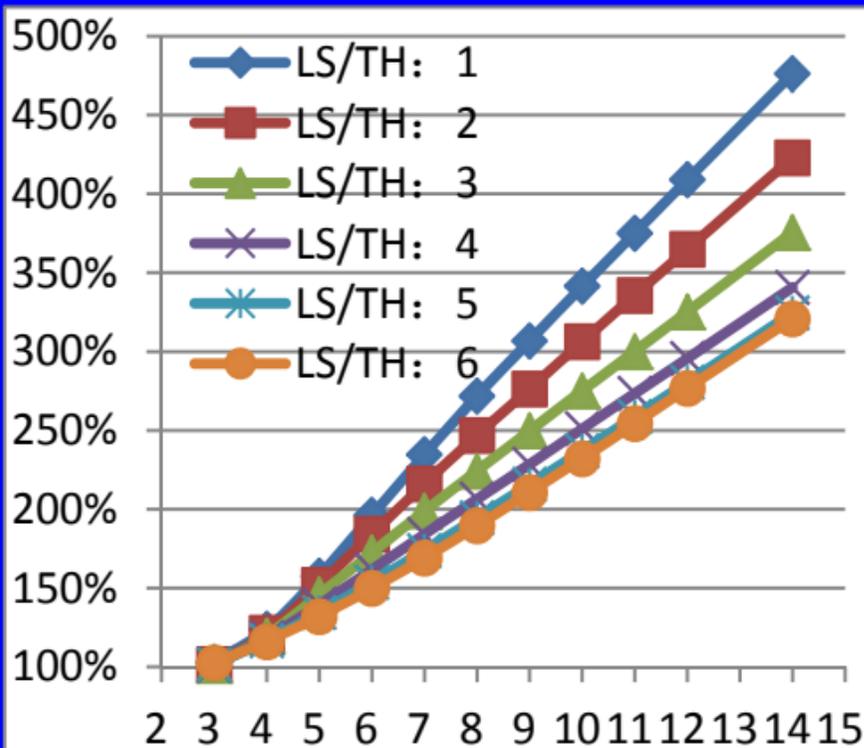
THM 保留了最小一乘的主要特征及优点。

## THM 反常情形的补充说明

对各组合的千组随机数计算，统计量  $|e_i|_M$  THM 比 LSM 大的反常出现在  $l = |e_{\max}/\sigma|$  接近 3 时。例如  $n=14$  且  $l=3$  时，单侧异常点位置有 7 种可能。在 7000 次回归中  $(|e_i|_{M_j})_{\text{THM}} / (|e_i|_{M_j})_{\text{LSM}} > 1.05$  并且  $|e_i|_M > 1.645\sigma$  的情形仅占 1.1%，占  $l=3 \sim 16$  时 84000 次回归的 0.08%。 $l=3$  时  $(|e_i|_{M_j})_{\text{THM}} / \sigma$  一般也小。例如异常点在第 4 点时， $(|e_i|_{M_j})_{\text{THM}} / (|e_i|_{M_j})_{\text{LSM}} > 1.05$  出现时均有  $(|e_i|_{M_j})_{\text{THM}} < 1.96\sigma$ ，说明“反常”对 THM 的稳健性影响不大。

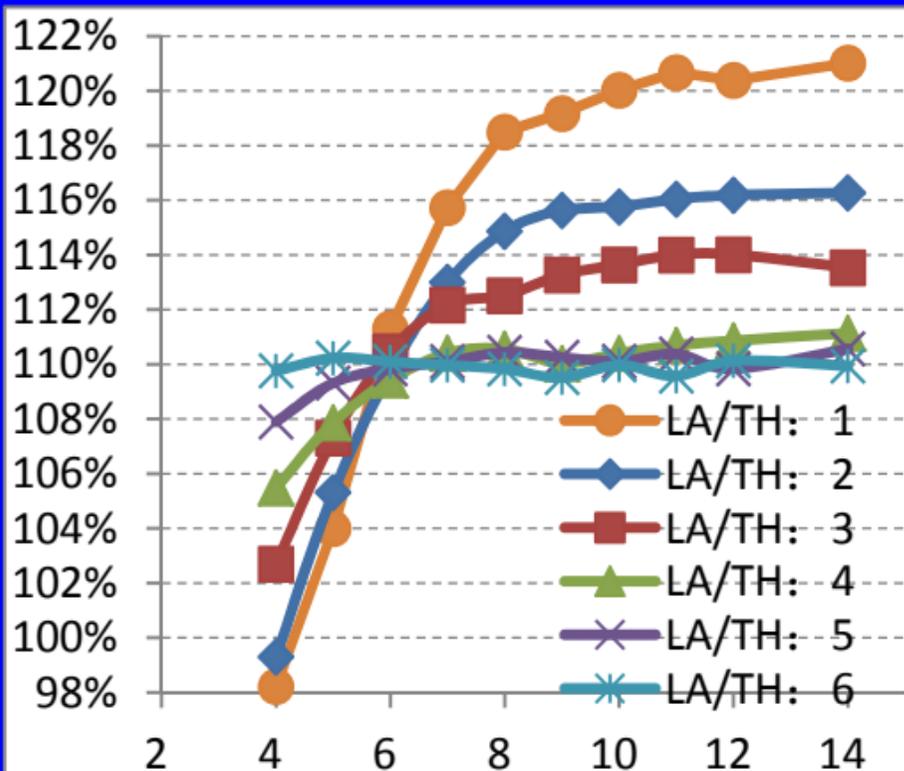
MCM 计算前对千组正态随机数按  $\chi^2$  分布及 t 分布作了截尾筛选，截去了概率小于 0.5% 或 1% 的部分。进一步的研究将增大  $m$  以优化 MCM 的收敛性，研究  $(|e_i|_{M_j})$  的分布规律。

图 1.  $n=12$  时粗差分别在第 1-6 点位  $\frac{(\overline{|e_{i|j}|})_{LSM}}{(\overline{|e_{i|j}|})_{THM}}$  的曲线



左图中横轴是  $l = |e_{\max}/\sigma|$  的值,  $n=12$ 。图中 6 条曲线表示粗差点分别在第 1-6 点时的情形。从图可定性看出：粗差点在一端时 THM 的稳健性更突出（最上两条）。

图 2.  $n=12$  时粗差分别在第 1-6 点位  $\frac{(\overline{|e_{i|j}|})_{LAM}}{(\overline{|e_{i|j}|})_{THM}}$  的曲线



左图中 6 条曲线表示粗差点分别在第 1-6 点时的情形. 从图可定性看出:

粗差点在一端时 THM 的稳健性更突出。

LAM 也有一定的稳健性。

l 愈大稳健性差别愈明显。

# 5. 实验教学中应用的案例

用最小偏向角法测三等角棱镜材料的色散曲线  $n_i = b_0 + b_1 \lambda_i^{-2.4}$ , 当分光计游标分度为  $1'$  时, 由一般方位角  $\varphi$ 、最小偏方位角  $\varphi_t$  的不确定度  $U_\varphi \approx 1.06'$ 、 $U_{\varphi_t} \approx 1.64'$ , 可得 ZF1、ZF2 玻璃折射率的扩展不确定度  $U_n \approx 6.7 \times 10^{-4}$ . [2] 往往有百分之几的同学会将方位角  $\varphi_t$  错读 1 大格, 粗差为  $\pm 30'$ , 对应  $n_i$  的偏差约  $e_{n \max} \approx \pm 5.1 \times 10^{-3} \approx 7.6 U_n$ . 用 He 光谱较亮的 7 谱线实验时, 6 对数据  $n_i$  的误差取  $\pm 6.7 \times 10^{-4}$  区间内均匀分布随机数, 最短波长处  $n_i$  的粗差取  $\pm 5.0 \times 10^{-3}$  (这时对直线影响最大). 对 MCM 算得的千组数据统计量, 结果的相对值见下表.

表 3. 色散曲线 7 波长直线化拟合时, 不同方法的稳健性比较

		最小二乘	最小一乘	技术综合
残差 $v_i$ 特征值	标准差 $s_y$ 的千组均值 $\overline{(s_y)_j}$	<u>100%</u>	121.7%	127.7%
	绝对值均值 $\overline{ v_i }$ 的千组均值 $\overline{\overline{ v_i }_j}$	134.0%	<u>100%</u>	102.4%
误差 $e_i$ 特征值	绝对值均值 $\overline{ e_i }$ 的千组均值 $\overline{\overline{ e_i }_j}$	438%	130%	<u>100%</u>
	最大绝对值 $ e _{\max}$ 千组均值 $\overline{ e _{\max j}}$	494%	131%	<u>100%</u>
	最大绝对值 $ e _{\max} \geq 1.2U_n$ 的比例	100%	12.6%	1.6%

从上表最后一行可看出, 即使某一角度错读了  $30'$ , 用稳健性好的技术综合法拟合, 在测量范围内仍然可得出合理色散曲线, 使预报值  $n_i$  的误差不大于  $1.2U_n$ . 从上表也可看出: 虽然 THM 的残差特征量与 LAM 相差小于 5%, 但误差特征量反映的 THM “抗差” 性较显著地优于 LAM, 更是大大优于 LSM.

# 附录3 表 4A. $n=12$ 时 SFM(含逐差法)与 LSM 的稳健性比较

统计量比较	$\left(\overline{ e_i _j}\right)_{\text{SFM}} / \left(\overline{ e_i _j}\right)_{\text{LSM}}$	$\left(\overline{ e_i _{Mj}}\right)_{\text{SFM}} / \left(\overline{ e_i _{Mj}}\right)_{\text{LSM}}$
异常值点位序号	$l=2.56, 3\sim 16$	$l=2.56, 3\sim 16$
1	84.5%	81.8%
2	95.1%	93.3%
3	106.3%	108.0%
4	116.3%	127.1%
5	123.1%	151.4%
6	126.1%	174.7%
平均	109%	122.7%

$n=12$  时表 4A 中的计算结果表明, 仅当粗差点靠近直线段端部时, SFM 的稳健性才比 LSM 略好, 但当粗差点不靠近端部时, SFM 的稳健性比 LSM 更差.

表 4B.  $n=12$  时 SFM(含逐差法)与 LSM 的  $(\overline{U_{b_1, A}})$  比较

统计量比较	$(\overline{U_{b_1, A}})_{\text{SFM}} / (\overline{U_{b_1, A}})_{\text{LSM}}$	$(\overline{U_{b_1, A}})_{\text{SFM}} / (\overline{U_{b_1, A}})_{\text{LSM}}$
异常值点位序号	$l =  e_{\max} / \sigma  = 2.58$	$l =  e_{\max} / \sigma  = 6$
1	133%	140%
2	131%	135%
3	129%	131%
4	130%	130%
5	127%	127%
6	129%	128%
平均	130.1%	131.7%

因自由度减半, SFM 的斜率 A 类不确定度比 LSM 平均大 30%, 这也是 SFM 在国内外学术界极少被运用的主因之一。

用 MCM 可证:  $n=8, m=2^{14}$  时 SFM 的斜率 A 类不确定度比 LSM 平均大 180%; 若直线过原点, 逐差法将比 LSM 大 12 倍!

## 哲学反思:

计算机普及前, 理论优位的科学哲学占主导地位, 实验统计学追求方法表述的公理化体系的完整性与普适性, 因为误差分布特征难以研究而只能讨论残差分布特征. 科学不一定首先看作是表征与观察世界的方式, 而更需重视科学能作为操作、介入世界的方式, 即是一种作用于世界的方式.<sup>[郭贵春等, 科学哲学的新进展, 200801 科学版]</sup> 现代科技中, 在根据目标任务而界定的广域内, 能用 MCM 来研究误差分布特征量、误差类统计量, 得出与概率相联系的直线拟合方法比较的相对结论.

## 参考文献

- [1] 陈希孺, 数理统计学小史等文, 连载于杂志《数理统计与管理》
- [2] George A.F. Seber, Alan J. Lee. Linear regression analysis (2nd ed.), Hoboken, N.J.: Wiley-Interscience, c2003
- [3] Kutner, Michael H. Applied linear regression models. Boston ; New York : McGraw-Hill/Irwin, c2004. 4th ed.
- [4] 朱鹤年, 新概念基础物理实验讲义[M], 清华大学出版社, 2013.
- [5] 纳特雷拉 M. 著, 毛镇道 等译. 实验统计学[M]. 上海: 上海翻译出版公司, 1990.
- [6] 陈希孺, 王松桂. 近代回归分析—原理方法及应用[M]. 合肥: 安徽教育出版社, 1987. P146.
- [7] 彭长清. 误差与回归[M], 兵器工业出版社, 1991, P129.
- [8] 国家标准: GB/T 3359-2009, 数据的统计处理和解释—统计容忍区间的确定