

实验误差分析中的概念及意义*

李潮锐¹, 郑碧华²

(1. 中山大学物理学系, 广东 广州 510275;
2. 湛江海洋大学物理学系, 广东 湛江 524025)

摘 要: 实验误差分析是实验教学中的一个重要环节, 但学生常因对误差概念理解不透彻而得出错误的实验结果。针对这一问题, 提出从实验测量的角度帮助学生理解误差统计规律, 并根据多年的实验教学经验, 对实验误差处理中的概念及意义作了分析。

关键词: 误差分析; 测量不确定度; 最小二乘法

中图分类号: O4-34 **文献标识码:** A **文章编号:** 0529-6579 (2003) S1-0150-04

实验误差是实验测量的客观结果, 实验教学过程必须在指导学生尽可能地减小系统误差和过失误差的基础上, 教导学生运用正确分析随机(偶然)误差的科学方法, 学习正确分析实验误差来源从而找出最佳的实验(观测)条件, 同时学会对实验结果的正确评估和描述。通过采用正确的数据处理分析, 从误差中发现实验问题, 提出实验改进方案。

根据定义, 误差即为“测量值与真值之差”, 真值是指某物理量客观存在的确定值。依此定义, 学生可能会提出这样的疑问: 如果已经知道被测对象的真值, 实验测量的目的是什么? 事实上, 真值通常是未知的。我们只能获得实验测量值, 那么误差分析的意义是什么? 教师必须使学生明白实验数据的处理与分析的目的是通过运用规范的实验数据(统计)分析方法, 从多次的测量数据中, 估算出最接近真值的实验测量结果及其可信度, 并探讨实验误差的可能来源。使学生认识到实验误差的存在客观性, 从而理解实验误差不能成为实验成败的判据。

随机误差(偶然误差)由某些不易控制的因素造成的, 但若对某一量值进行足够多次的等精度测量, 就会发现随机误差服从统计规律(高斯分布或正态分布), 误差的出现完全由概率决定。实验数据的处理与分析探讨如何降低随机误差的最有效方法。可以通过设计适当的实验测量对象, 并对其进行大量重复测量分析(在相同的实验条件下), 使学生对测量误差的统计分布规律有更直观的认识。

1 平均值和标准偏差的意义

实验误差分析是本科实验绪论课的主要内容, 一方面, 由于低年级学生对概率统计了解不多, 从而实验测量值分布规律遵从高斯函数的“理论解释”是教学难点之一; 另一方面, 正确认识、理解实验测量的高斯分布是正确运用实验数据处理方法及合理的实验结果描述的基础。为此, 实验课程教师都想方设法采用“生动”“直观”的理论语言对这一概念作出“最清晰”的解释^[1]。事实上, 教师应该尝试在实验课程教学过程中有意识地引导学生从实验的角度认识规律, 这本也应该是实验课程的教学目的之一。不要过分依赖于理论推倒, 特别是对一些首次出现的关键性概念或对实验现象的观测。

为了解决上述教学问题, 教师可以从硬件和软件两方面寻找教具。硬件上, 可以选择“G-M计数及统计分析”实验项目为实验误差理论分析的“课件”, 通过采用自动控制数据采集并对实验数据进行实时处理显示。通过这一实验演示, 配合教师课堂上讲解, 帮助学生直观地理解实验测量的统计规律, 同时也使学生对实验课程的学习方法有了初步的认识。软件方面, 采用计算机模拟方法, 编程(或直接使用专业软件平台)模拟物理量测量过程, 并对“测量”结果实时统计处理、显示。总之, 实验课程的教学问题尽量通过实验操作来解决。

在认识测量统计分布规律的基础上, 更为重要的是理解平均值 $\bar{x}$ 和标准偏差 σ 的意义。从图1的

* 收稿日期: 2003-05-09

基金项目: 中山大学教学改革研究课题资助项目

作者简介: 李潮锐(1962年生), 男, 副教授; E-mail: crli@gzsums.edu.cn

“测量”结果可得:测量平均值 $\bar{x}$ 是最可几实验值, 越偏离 $\bar{x}$ 的测量值出现的几率越小。从而也说明:测量偏差是实验测量的客观事实。通过简单的数学运算可得出:测量值出现在 $\bar{x} + \sigma > x > \bar{x} - \sigma$ 范围内的几率为 68.3%; $\bar{x} + 2\sigma > x > \bar{x} - 2\sigma$ 范围内的几率为 95.4%; $\bar{x} + 3\sigma > x > \bar{x} - 3\sigma$ 范围内的几率为 99.7% 等等。通常, 为提高实验测量精度, 实验教科书还介绍多种剔除误差较大实验值的判断方法。对于这一“技术”问题, 教师应帮助学生正确理解、运用这些方法, 防止学生产生“实验误差是衡量实验成败的判据”的错误观点。

等条件独立多次重复测量的平均值 $\bar{x}$ 和标准偏差 σ 可分别表示为:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2 \quad (2)$$

我们可以发现:当 $n=1$ 时, $\bar{x} = x_1, \sigma = 0$! 即, 当只有一次测量时, 平均值为一次测量值, 且偏差为零。其意为:一次测量具有最高精密性! 显然, 这违反了测量误差的客观事实! 学生要问:为什么会出现这一“谬论”? 回答这一问题的关键在于随机误差符合高斯分布的前提条件:每单次独立测量必须要求在相同的环境下进行, 且互不相关。事实上, 当考虑有限次测量时, 特别是在极端情况, 测量值的相关性显然不可忽视。根据定义, 当 $n \rightarrow \infty$ 次测量时, 标准偏差为

$$\sigma_n = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \quad (3)$$

其中, n 为等条件独立测量次数。对于有限 n 次测量, 平均值等式使独立变量减少一个, 即等条件独立测量次数只有 $n-1$, 从而可定义有限 n 次测量标准偏差为

$$\sigma = \sigma_{n-1} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (4)$$

当 $n \rightarrow \infty$ 时, 分母为 n 或 $n-1$ 是没有差别, 即 σ_n 与 σ_{n-1} 近似相等。但是, 当 $n=1$ 时, σ_{n-1} 没意义, 也就是说, 只有一次测量无法确定测量的标准偏差^[2-4]。从定义 σ_n 到 σ_{n-1} 引入的过程, 没有严格的理论推导, 我们只需要理解“等条件独立测量”的物理含意, 通过思维推理便能得到合理分析结果。

以上定义的标准偏差 (σ_n 或 σ_{n-1}) 仅代表所有测量数据与平均值之间平均的偏差量, 也暗示在 n 次

测量中的每一次测量数据均有 68.3% 的几率出现在 $[\bar{x} - \sigma, \bar{x} + \sigma]$ 范围内。但事实上, 我们关心的是测量平均值 (或最佳期望值) 的可信度, 而非单次测量的可能偏差 (误差)。

2 平均值标准偏差的意义

通常一个物理量是经由测量数个其他物理量, 再藉由它们之间的物理关系计算而得出。例如: 动量是由质量与速度测量值相乘而得 (速度又由位移与时间测量值得出)。当测量时, 质量、位移与时间的个别误差将影响最后结果的误差。这是误差传递问题。

假设 x 代表某一个物理量, 并由 $(u, v, \cdots)$ 等测量值所决定, 即 $x = f(u, v, \cdots)$ 。若以 $(\bar{u}, \bar{v}, \cdots)$ 分别代表 $(u, v, \cdots)$ 等量的平均值, 则平均值 $\bar{x} = f(\bar{u}, \bar{v}, \cdots)$ 。

对于某一组测量样本数据, 可以表示为 $x_i = f(u_i, v_i, \cdots)$ 则

$$\begin{aligned} x_i - \bar{x} &= f(u, v, \cdots) - f(\bar{u}, \bar{v}, \cdots) = \\ &\left(\frac{\partial x}{\partial u}\right)_{\bar{u}} (u_i - \bar{u}) + \left(\frac{\partial x}{\partial v}\right)_{\bar{v}} (v_i - \bar{v}) + \cdots \end{aligned} \quad (5)$$

(一级近似展开)

测量值的方差

$$\begin{aligned} \sigma_x^2 &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \\ &\sigma_u^2 \left(\frac{\partial x}{\partial u}\right)_{\bar{u}}^2 + \sigma_v^2 \left(\frac{\partial x}{\partial v}\right)_{\bar{v}}^2 + \sigma_{uv}^2 \left(\frac{\partial x}{\partial u}\right)_{\bar{u}} \left(\frac{\partial x}{\partial v}\right)_{\bar{v}} + \cdots \end{aligned} \quad (6)$$

其中,

$$\begin{aligned} \sigma_u^2 &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (u_i - \bar{u})^2, \sigma_v^2 = \\ &\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (v_i - \bar{v})^2, \\ \sigma_{uv}^2 &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (u_i - \bar{u})(v_i - \bar{v}) \end{aligned}$$

如果 u 和 v (被测物理量) 彼此不相关, 则 σ_{uv}^2 为零。通常, 绝大多数实验测量的各个参数间是互不相关的^[2]。于是方差可以简化为

$$\sigma_x^2 \approx \sigma_u^2 \left(\frac{\partial x}{\partial u}\right)_{\bar{u}}^2 + \sigma_v^2 \left(\frac{\partial x}{\partial v}\right)_{\bar{v}}^2 + \cdots \quad (7)$$

根据平均值定义式 (1) 可知, 平均值是由各独立测量值取平均而得到, 即可理解为: 平均值是以各测量值为独立变量的函数。从误差传递的角度, 平均值误差来源于独立测量值 $(x_1, x_2, \cdots, x_n)$ 的误差, 且平均值标准偏差 $\sigma_{\bar{x}}$ 由式 (7) 决定, 即

$$\sigma_{\bar{x}}^2 = \sum_{i=1}^n \left[\sigma_{x_i}^2 \left(\frac{\partial \bar{x}}{\partial x_i}\right)^2 \right] = \frac{\sigma^2}{n} \quad (8)$$

即平均值标准偏差 $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

其意义是:每次(组)多次实验所得平均值都不会相同,这些平均值也会形成一种分布,平均值标准偏差便是代表这些不同的平均值的可能差异性(精密度)。

从上述的分析结果,可得出结论:标准偏差 σ 显示单次实验测量值与平均值间可能偏差的程度(单次测量的精密度);重复(增加实验次数)并不会减少其数值。平均值标准偏差 $\sigma_{\bar{x}}$ 则显示所得平均值的可重复性程度(即,结果的精密度);多组重复测量所计算出平均值的标准差,其数值可以通过增加测量次数而减少,且与 $\sqrt{n}$ 成反比。

3 曲线拟合分析及应用

事实上,我们常常需要测量某物理量(应变量)随物理参数(自变量)变化情况,分析它们彼此间的关系,从而研究变化过程的内在联系。例如:电阻(纵坐标)随温度(横坐标)的变化关系。通过曲线拟合分析进一步强化学生的实验数据准确定量分析能力,引导学生尽可能避免使用凭直觉的“描点画线”“技巧”。最小二乘法^[2,5]便是寻找出一条合理的描绘实验数据关系曲线(或函数,例如电阻-温度函数关系)的分析方法,它使所有数据点到曲线距离的平方总和(方差)为最小。与标准偏差定义式(2)比较,我们可以发现,这一方法的实质是使拟合点(或最可几测量值)与实验测量值偏差最小。

若用 (x_i, y_i) 表示实验数据点,即需要找出 $y = f(x)$,使得

$$\delta = \sum_{i=1}^n [f(x_i) - y_i]^2 \quad (9)$$

为最小值。注意:以上假设自变量没有误差,或相对很小可忽略。

应用例1:直线拟合

拟合分析的最简单例子是应变量和自变量之间为线性关系。通过实验数据点 (x_i, y_i) , 拟合关系 $y = ax + b$, 希望找出参数 (a, b) 使拟合曲线满足

$$\delta = \sum_{i=1}^n [f(x_i) - y_i]^2 = \sum_{i=1}^n (ax_i + b - y_i)^2$$

为最小值。

由 $\frac{\partial \delta}{\partial a} = 0$ 和 $\frac{\partial \delta}{\partial b} = 0$, 可得到

$$a = \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i - n \sum_{i=1}^n x_i y_i}{\left(\sum_{i=1}^n x_i\right)^2 - n \sum_{i=1}^n x_i^2},$$

$$b = \frac{\sum_{i=1}^n x_i y_i \sum_{i=1}^n x_i - \sum_{i=1}^n y_i \sum_{i=1}^n x_i}{\left(\sum_{i=1}^n x_i\right)^2 - n \sum_{i=1}^n x_i^2}$$

上式中 a 为直线斜率, b 为其截距。根据误差传递,可求出 a 和 b 的误差。认为 a, b 为 x_i 和 y_i 函数,并假设 x_i 没有误差, a, b 的误差是由 y_i 的误差所传递。

$$\Delta = \left(\sum_{i=1}^n x_i\right)^2 - n \sum_{i=1}^n x_i^2$$

$$\text{则 } \frac{\partial a}{\partial y_i} = \frac{1}{\Delta} \left(\sum_{i=1}^n x_i - nx_i\right),$$

$$\text{且 } \frac{\partial b}{\partial y_i} = \frac{1}{\Delta} \left(x_i \sum_{i=1}^n x_i - \sum_{i=1}^n x_i^2\right)$$

$$\text{可得到 } \sigma_a^2 = -\frac{n\sigma^2}{\Delta}, \quad \sigma_b^2 = -\frac{\sigma^2}{\Delta} \sum_{i=1}^n x_i$$

其中, $\sigma^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - ax_i - b)^2$ 为曲线拟合方差。由于极小值条件限制使独立测量实验点的数目减2,从而上式中因子为 $n-2$ 而非 n 。线性相关系数为

$$\gamma = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \sqrt{n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i\right)^2}}$$

γ 的大小反映了拟合函数与实验点的近似程度,若 γ 的绝对值越接近于1则表示 $x \sim y$ 数据间越适合用上述线性关系描述。 γ 对所拟合曲线的合理性给出一个定量的说明^[2,6]。

应用例2:有限次测量实验平均值是最佳期望(估算)值。

平均值 $\bar{x}$ 是实验测量值,即独立变量 $(x_1, x_2, \dots, x_n)$ 的函数,它们的关系由式(1)决定

根据最小二乘法,要求 $\bar{x}$ 的取值使

$$\delta = \sum_{i=1}^n [f(x_i) - y_i]^2 = \sum_{i=1}^n (x_i - \bar{x})^2$$

为极小,即

$$\frac{\partial \delta}{\partial \bar{x}} = \frac{\partial}{\partial \bar{x}} \left(\sum_{i=1}^n (x_i - \bar{x})^2\right) = -2 \sum_{i=1}^n x_i + 2n\bar{x} = 0$$

可得到 $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$, 有限次测量实验平均值是最佳期望(估算)值。

参考文献:

- [1] 郑裕芳, 李仲荣. 近代物理实验[M]. 广州: 中山大学出版社, 1989.
- [2] Bevington P R, Robinson D K. Data reduction and error analysis for the physical sciences[M]. 3rd edi. McGraw-Hill Science/Engineering/Math, 2002.
- [3] Taylor J R. Introduction to error analysis: The study of uncertainties in physical measurements[M]. University Science Books, 1982.
- [4] Meyer S L. Data analysis for scientists and engineers[M]. Peer Management Consultants Ltd, 1992.
- [5] Mandel J. The statistical analysis of experimental data[M]. Dover Pubns, 1984.
- [6] Rabinovich S, Alferieff M E. Measurement errors and uncertainties: theory and practice[M]. 2nd edi. Amer Inst of Physics, 1999.

Error Analysis in Physical Measurements

LI Chao-rui¹, ZHENG Bi-hua²

(1. Department of Physics, Sun Yat-sen University, Guangzhou 510275, China;

2. Department of Physics, Zhanjiang Ocean University, Zhanjiang 524025, China)

Abstract: Error analysis is a minor part of the course in experimental physics, but it is too often short-changed so that students recite its equations without understanding their rationale. Building on the fundamentals of measurement theory, an introduction to the concepts of error analysis in physical measurements is provided, also the methods of estimating measurement errors and uncertainty is presented for serious experimentaists.

Key words: error analysis; measurement uncertainty; method of least squares