

磁屏蔽材料、高性能磁头材料、高磁弹性材料等方面都获得了可喜的研究成果,许多已制成可应用的器件。可以相信,继续抓紧非晶态磁性的研究,增强基础研究和应用研究的联系和配合,则在不久的将来,非晶态磁性材料将会在我国国民经济中发挥出愈来愈大的作用。

参 考 文 献

- [1] P. Duwez, R. H. Willems and W. Klement, *J. Appl. Phys.*, **31** (1960), 1136.
 [2] A. I. Gubanov, *Sov. Phys. Solid St.*, **2** (1960), 468.
 [3] F. E. Luborsky and L. A. Johnson, *J. de Phys.*, **41** (1980), C8—820.
 [4] 王荫君、赵见高, *物理*, **7**(1978), 24.
 [5] W. S. Chan, B. G. Shen, H. Y. Lo and B. L. Yu, in *Proc. 4th Int. Conf. on Rapidly Quenched Metals*, (1981), 1137.
 [6] S. J. Poon and J. Durand, *Phys. Rev. B*, **18** (1978), 6253.
 [7] 郑德娟、王震西、张殿琳、林淑媛, *物理学报*, **31** (1982), 185.
 [8] C. D. Graham and T. Egami, *Ann. Rev. Mater. Sci.*, **8** (1978), 423.
 [9] J. P. Rebouillat, A. Lienard, J. M. D. Coey, R. A. Boggiano and J. Chappert, *Physica*, **B+C** **86—88** (1977), 773.
 [10] J. M. D. Coey, J. Chappert, J. P. Robouillat and T. S. Wang *Phys. Rev. Lett.*, **36**(1976), 1061.
 [11] A. Aharony and Pytte, *Phys. Rev. Lett.*, **45** (1980), 1583.
 [12] M. J. O'shea and D. J. Sellmyer, *J. Appl. Phys.*, **53** (1982), 7722.
 [13] H. A. Mook, *J. Appl. Phys.*, **49** (1978), 1665.
 [14] Y. Takanashi and M. Shimizu, *Phys. Lett.*, **A** **58** (1976), 419.
 [15] T. Egami, *Mat. Res. Bull.*, **13** (1978), 557.
 [16] K. Moorjani, S. K. Ghatak, K. V. Rao, B. Kramer and H. S. Chem, *J. de Phys.*, **41** (1980), C8—718.
 [17] J. W. Lynn, R. W. Erwin, J. J. Rhyne and H. S. Chen, *J. Appl. Phys.*, **52** (1981), 1738.
 [18] 杨翠英、王荫君、王忠铨、李方华, *金属学报*, **15** (1979), 351.
 [19] R. Alben, J. Becker and M. Chi, *J. Appl. Phys.*, **49** (1978), 1653.
 [20] S. G. Conelison, D. J. Sellmyer, J. G. Zhao and Z. D. Chen, *J. Appl. Phys.*, **53** (1982), 2330.

误差理论与实验的数学处理讲座

第三讲 测量误差与实验结果的处理

高 崇 寿

(北京大学物理系)

一、样本和统计量

如果测量某一物理量 x , 独立测量了 N 次得到了 N 个数值 $x_i, i = 1, \dots, N$. 显然这些 x_i 是随机变量 x 的 N 个随机取值, 这 N 个 x_i 的整体从概率意义上应当反映随机变量 x 的概率分布特征. 当然, 由于 N 是有限数, 此“反映”不可能是完全的. 这些 x_i 的集合称为随机变量 x 的一个随机样本, N 称为样本的容量. 实验中首先要处理的是如何从随机变量的样本中来得得到有关这个随机变量所代表的物理量的信息.

容量为 N 的样本 $x_i, i = 1, \dots, N$, 是一个随机变量的 N 个随机取值. 同时, 我们又可以把 $x_i (i = 1, \dots, N)$ 看作是 N 个随机变量的一组随机取值. 为了便于区分, 我们把这 N 个随机变量仍记作为 $x_i, i = 1, \dots, N$, 而把这 N 个随机变量的一组具体随机取值用 $x_i^{(1)}, i = 1, \dots, N$, 代表. 这样另一组随机取值可记作 $x_i^{(2)}, i = 1, \dots, N$, 依此类推.

在这种考虑下, 这 N 个随机变量是互相独立的. 并且由于它们实际上都反映用同一仪器和相同方法测量同一物理量 x , 因此它们的概率密度函数形式应完全相同, 即都是 x 的概率密度函数, 只不过把自变量换成相应的 x_i . 这

样它们的联合概率密度函数为

$$p(x_1, \dots, x_N) = \prod_{i=1}^N p(x_i). \quad (3.1)$$

在对实验结果进行处理时,常常需要处理随机变量 $x_i (i = 1, \dots, N)$ 的某种函数 $y = y(x_i)$, 这样的函数称为统计量. 统计量是根据实际实验分析工作的需要而选取的. 例如,

样本平均值 $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$ 和样本方差

$$S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$$

就是最常用的统计量.

如前所述,各 x_i 是 x 的 N 次独立测量值,并且这 N 次测量所采用的仪器和方法都是相同的,这就表明这 N 个测量值 $x_i (i = 1, \dots, N)$ 对物理量 x 提供信息的重要性也是相同的. 如果选用了对各 x_i 不是完全对称的函数形式作为统计量来处理实验的结果,这就表明实际上对某些次测量过分看重了,这种偏向带有人为的特点,显然对于客观地得到实验所提供的信息是不利的. 因此,在实际工作中一般总是选取对各 x_i 是完全对称的函数作为统计量. 这个完全对称性表现为: 将任意两个 x_i 和 $x_j (i \neq j)$ 互换,统计量的值完全不变.

统计量的完全对称性质在实际计算时非常有用,特别是在求统计量的期待值时,利用这个对称性质可以使计算大大简化. 由于各 x_i 是互相独立的,它们的联合概率密度函数是各 x_i 的概率密度函数的乘积,对不同 x_i 求期待值的运算实际上是独立进行的. 这个性质在统计量是 x_i 的多项式时特别有用. 以样本方差的期待值为例,

$$S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2,$$

由于 \bar{x} 也是 x_i 的一次齐次函数,因此 S_x^2 是各 x_i 的二次齐次函数. 在求期待值后,各 x_i 变量均不会再出现,并且由于 $\langle x_i^n \rangle = \langle x^n \rangle$, $\langle S_x^2 \rangle$ 一定是 $\langle x^2 \rangle$ 和 $\langle x \rangle^2$ 的线性组合. 具体计算如下:

物理

$$\langle S_x^2 \rangle = \frac{1}{N-1} \sum_{i=1}^N \langle (x_i - \bar{x})^2 \rangle$$

$$= \frac{N}{N-1} (\langle x^2 \rangle - 2\langle x_i \bar{x} \rangle + \langle \bar{x}^2 \rangle),$$

考虑到

$$\langle x_i \bar{x} \rangle = \frac{1}{N} \sum_{j=1}^N \langle x_i x_j \rangle = \frac{1}{N} [\langle x^2 \rangle$$

$$+ (N-1)\langle x \rangle^2],$$

$$\langle \bar{x}^2 \rangle = \frac{1}{N^2} \sum_{i,j=1}^N \langle x_i x_j \rangle = \frac{1}{N^2} [N\langle x^2 \rangle$$

$$+ N(N-1)\langle x \rangle^2]$$

$$= \frac{1}{N} [\langle x^2 \rangle + (N-1)\langle x \rangle^2],$$

代入得

$$\langle S_x^2 \rangle = \langle x^2 \rangle - \langle x \rangle^2 = \sigma^2(x). \quad (3.2)$$

再例如计算统计量 $\sum_{i=1}^N \sum_{j=1}^N (x_i - x_j)^2$ 的期

待值

$$\left\langle \sum_{i=1}^N \sum_{j=1}^N (x_i - x_j)^2 \right\rangle = N(N-1) \langle (x_i - x_j)^2 \rangle$$

$$= 2N(N-1) (\langle x^2 \rangle - \langle x \rangle^2)$$

$$= 2N(N-1)\sigma^2(x).$$

在有关统计量的各种计算中,除了经常需要计算各次矩(即随机变量各幂次的期待值)之外,有时还需要直接知道统计量的概率密度函数. 根据第二讲给出的(2.27)式,统计量 $y = f(x_i)$ 的概率密度函数 $p(y)$ 由

$$p(y) = \int \cdots \int \delta(y - f(x_1, \dots, x_N)) \cdot$$

$$\prod_{i=1}^N [p(x_i) dx_i]$$

给出. 实际作时可以先对 x_1 积分,去掉 δ 函数,然后再积分 x_2, \dots, x_N , 积分时要注意利用 x_2, \dots, x_N 的完全对称性.

还可以用其它方法求统计量的概率密度函数,在具体问题中常常能根据实际情况找到特殊的简单办法.

二、直接测量的误差处理

现在考察测量一个物理量 x 的情形, 这时真值 x 是确定的但却是未知的。在实验中如果已消除了系统误差, 则由于统计误差的存在, 测得的观测值应在真值附近摆动, 从而表现为一定的概率分布, 真值 x 应是这个随机变量的期待值。按定义, 期待值应是无穷多观测值的平均值, 但是实际上不可能进行无穷多次测量, 只能在一定条件下测量次数尽可能多。若实际进行了 N 次测量, 则得到一个容量为 N 的样本 x_i , $i = 1, \dots, N$ 。

如果测量的标准误差已知为 $\sigma(x)$, 则每一次测量的结果表述为

$$x = x_i \pm \sigma(x). \quad (3.3)$$

该式的含义是对物理量 x 测量给出的估值是 x_i , 而 x_i 的概率分布的标准误差为 $\sigma(x)$ 。如果随机变量 x_i 服从的是正态分布, 则上式的含义为: x 的真值在 $x_i - \sigma(x)$ 到 $x_i + \sigma(x)$ 之间的概率是 68.3%。

如果进行了 N 次测量, 可以取样本平均值作为物理量的估值, 因为样本平均值 \bar{x} 的期待值等于随机变量 x_i 的期待值, 即

$$\begin{aligned} \langle \bar{x} \rangle &= \left\langle \frac{1}{N} \sum_{i=1}^N x_i \right\rangle = \frac{1}{N} \sum_{i=1}^N \langle x_i \rangle \\ &= \langle x \rangle. \end{aligned} \quad (3.4)$$

但是其方差为

$$\begin{aligned} \sigma^2(\bar{x}) &= \langle \bar{x}^2 \rangle - \langle \bar{x} \rangle^2 \\ &= \left\langle \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N x_i x_j \right\rangle - \langle x \rangle^2 \\ &= \frac{1}{N^2} \left(\sum_{i=1}^N \langle x_i^2 \rangle + \sum_{\substack{i,j=1 \\ i \neq j}}^N \langle x_i x_j \rangle \right) - \langle x \rangle^2 \\ &= \frac{1}{N^2} [N \langle x^2 \rangle - N(N-1) \langle x^2 \rangle] - \langle x \rangle^2 \\ &= \frac{1}{N} \sigma^2(x), \end{aligned} \quad (3.5)$$

即 \bar{x} 的标准误差比 x_i 的标准误差减少到 $\frac{1}{\sqrt{N}}$,

实验结果可表述为

$$x = \bar{x} \pm \frac{1}{\sqrt{N}} \sigma(x). \quad (3.6)$$

这个结果表明在实验中进行多次测量对提高测量的精度是有意义的。

值得注意的是, 如果随机变量 x_i 服从正态分布, 则 \bar{x} 也服从正态分布。这时, 上式的含义是, 物理量 x 的真值在 $\bar{x} - \frac{1}{\sqrt{N}} \sigma(x)$ 到 $\bar{x} + \frac{1}{\sqrt{N}} \sigma(x)$ 之间的概率为 68.3%。如果并不知道 x_i 服从的是否是正态分布, 则 \bar{x} 服从的不一定是正态分布, 但根据本讲座第二讲中给出的定理二, 当 N 很大时 \bar{x} 渐近地服从正态分布。因此实际上只要测量次数比较多, 就可以当作正态分布来处理。

下面讨论两个问题:

1. 如何判断随机变量是否服从正态分布?

如果事先并不知道随机变量服从什么分布, 可以利用本讲座第二讲中的定理一来判定是否服从正态分布。具体作法是先把测量时可能出现的系统误差设法消除掉, 使测量时遇到的误差都属于统计误差的性质, 然后分析导致统计误差的各因素中是否有某因素起决定作用。如果有, 则应分析该因素导致的概率分布是什么, 由于该因素起决定作用, 这个分布也就决定了该物理量测量值服从的分布; 如果没有哪一个因素起决定因素, 亦即该物理量的统计误差是许多个互相独立的因素影响的综合效果, 而每一因素的影响都不起主要作用, 那末由定理一就可知道该物理量的观测值应服从正态分布。

由此可见, 在实验中只要把系统误差扣除, 并使测量的统计误差并不主要是由某个或少数几个因素决定, 则实验中得到的观测值实际上渐近地服从正态分布。

2. 如果测量的标准误差不知道, 怎么办? 最重要的办法是对某确定的物理量重复进行多次测量, 以确定其标准误差。实际上, 即使是已标明标准误差的仪器设备, 在用来作重要的精密实验之前, 也应进行上述核对标准误差的测量,

以便检验经过标定的标准误差是否正确。

在许多情况下，事先不知道测量的标准误差，并且也没有条件为确定标准误差而进行大量测量。换言之，一切信息的来源都是样本，在这种情况下测量的结果应如何处理？显然，仍可采用样本平均值为物理量的估值，问题在于如何确定测量误差的报道。下面分几步来进行讨论：

(1) 首先需要找一个统计量，其期待值为样本平均值的方差 $\frac{1}{N} \sigma^2(x)$ 。前面已讨论过，样本方差 $S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$ 的期待值是 $\sigma^2(x)$ 。正因为如此， S_x 称为样本的标准偏差，并常被作为标准误差的估计值。由此可以定义样本平均值的标准偏差 $S_{\bar{x}} = \frac{1}{\sqrt{N}} S_x$ ，作为样本平均值的标准误差 $\frac{1}{\sqrt{N}} \sigma(x)$ 的估计值。

(2) 但是，如果对测量结果采用如下的报道 $x = \bar{x} \pm S_{\bar{x}}$

$$= \bar{x} \pm \sqrt{\frac{1}{N(N-1)} \sum_{i=1}^N (x_i - \bar{x})^2}$$

仍是有问题的，因为按习惯自然会把这样的写法理解为 $S_{\bar{x}}$ 是 \bar{x} 的标准误差， x 在 $\bar{x} - S_{\bar{x}}$ 到 $\bar{x} + S_{\bar{x}}$ 之间的概率是 68.3%。但是实际上 $S_{\bar{x}}$ 虽然有确定的值，但它本身是一个随机变量。 $S_{\bar{x}}^2$ 的期待值 $\langle S_{\bar{x}}^2 \rangle$ 等于 $\frac{1}{N} \sigma^2(x)$ ，但 $S_{\bar{x}}^2$ 并不等于 $\frac{1}{N} \sigma^2(x)$ ，用 $S_{\bar{x}}$ 来代替 $\frac{1}{\sqrt{N}} \sigma(x)$ 时，对概率的估计就可能有偏离。此外，由于 $S_{\bar{x}}$ 本身是一个随机变量，就使原来的正态分布本身也受到影响到。

(3) 通常采用的观测结果的报道形式为 $x = x_i \pm \sigma(x)$ ，其前提是观测值 x_i 服从正态分布，或者说 $\frac{x_i - x}{\sigma(x)}$ 服从标准正态分布。现在即使样本平均值 \bar{x} 服从正态分布，但其标准误差 $\frac{1}{\sqrt{N}} \sigma(x)$ 未知，用 $S_{\bar{x}}$ 代替后所构成的

$\frac{\bar{x} - x}{S_{\bar{x}}}$ 并不服从标准正态分布。可以证明它满

足自由度为 $N-1$ 的 t 分布，因此对实验结果的报道应该从 t 分布的性质出发。

在 $N \rightarrow \infty$ 时， t 分布趋于标准正态分布，但是在 N 不很大时， t 分布远比标准正态分布要宽。

(4) 为了和通常习惯的对实验结果的报道方法取得一致，常常仍然按置信水平 68.3% 来标出误差范围。按照 t 分布

$$P, \left(\left| \frac{\bar{x} - x}{S_{\bar{x}}} \right| < t_{0.683} \right) = 0.683$$

来给出置信水平为 68.3% 的报道形式：

$$x = \bar{x} \pm t_{0.683} S_{\bar{x}}, \quad (3.7)$$

其中 $t_{0.683}$ 的数值由自由度数决定。在自由度数 $\nu = N-1 \rightarrow \infty$ 时， $t_{0.683} = 1.00$ ，在 ν 愈小时， $t_{0.683}$ 愈大。在 $\nu = 1$ 时， $t_{0.683} = 1.84$ ；在 $\nu = 2$ 时， $t_{0.683} = 1.32$ ；在 $\nu \geq 3$ 时，近似有 $t_{0.683} = 1 + \frac{0.58}{\nu}$ 。（准确到小数点后两位）。

这样给出的 (3.7) 式虽然也把置信水平取作 68.3%，但是这并不反映 $t_{0.683} S_{\bar{x}}$ 是正态分布随机变量 \bar{x} 的标准误差 $\sigma(\bar{x})$ 。虽然 (3.7) 式和 $x = \bar{x} \pm \sigma(\bar{x})$ 的置信水平是相同的，但含义完全不同。 $x = \bar{x} \pm \sigma(\bar{x})$ 是按正态分布给出的报道，(3.7) 式则是按 t 分布给出的报道。按 t 分布时，实际分布比正态分布要宽。正因为如此，在按 t 分布报道实验结果时，应注明自由度数，或采取高置信水平的误差报道。

在直接测量结果的数据处理中，另一项重要工作是进行样本 χ^2 检验。上面说到 $\frac{\bar{x} - x}{S_{\bar{x}}}$

服从自由度为 $N-1$ 的 t 分布，其根据是因为

样本 χ^2 量 $\frac{1}{\sigma^2} \sum_{i=1}^N (x_i - \bar{x})^2$ 服从自由度为 $N-1$

的 χ^2 分布，并且与样本平均值 \bar{x} 互相独立。这样，根据 χ^2 分布的性质，

$$\left\langle \frac{1}{\sigma^2} \sum_{i=1}^N (x_i - \bar{x})^2 \right\rangle = N-1, \quad (3.8)$$

要求 $\frac{1}{\sigma^2} \sum_{i=1}^N (x_i - \bar{x})^2$ 应在 $N-1$ 附近摆动。在 σ^2 未知时, 可以利用 (3.8) 式来给出 σ^2 的估值; 在 σ^2 是已知时, (3.8) 式一方面可以作为 σ^2 的一个检验, 另一方面如果已知 σ^2 的值是可靠的, 则可以用来考察这些样本是否互相协调。具体来说, 如果根据实验得到的样本 $x_i (i=1, \dots, N)$ 算出的样本 χ^2 量的值远远大于 $N-1$, 则应考虑以下几种可能性:

- (1) 已经给出的标准误差 σ 实际上偏小了;
- (2) 测量时有未扣除的系统误差;
- (3) 随机变量的分布不是正态分布;
- (4) 样本 χ^2 量远大于 $N-1$ 也有可能就是由统计涨落引起的。

要强调的是, 由于样本 χ^2 检验并不需要额外再进行测量, 完全是从已有的样本中就可以进行的了, 因此在对直接测量的结果进行处理分析时, 应该把样本 χ^2 检验作为其中一个重要的组成部分。

现在看一个实际例子: 用数字电压表测量一个 1.5 伏左右的电源电压。数字电压表用 2 伏的挡, 读数在小数点后有 4 位, 最后一位只显示 0 或 5。这种数字电压表的误差分布是均匀分布, 即最后一位数字离 0 近就跳到 0, 离 5 近就跳到 5。这个均匀分布的标识宽度的量 $\Delta = 0.00025$ 伏, 而数字电压表的标准误差为 $\sigma = \frac{1}{\sqrt{3}} \Delta = 0.00014$ 伏。但要注意的是, 这时并不是正态分布而是均匀分布。

对这个电源进行了 10 次测量, 结果如表 1 所示。这样给出电源电压的样本平均值为

$$\bar{v} = 1.51555 \text{ 伏。}$$

样本平均值的标准偏差为

$$S_{\bar{v}} = \sqrt{\frac{3.397 \times 10^{-4}}{90}} \text{ 伏} = 1.943 \times 10^{-3} \text{ 伏。}$$

考虑到现在的测量次数 $N=10$, 还应该用 t 分布来对测量结果给出报道。查出自由度 $\nu = 10 - 1 = 9$ 的 $t_{0.683} = 1.06, t_{0.90} = 31.833$, 给出置信水平 68.3% 的误差限为

表 1

i	v_i (伏)	$(v_i - \bar{v})^2$ (伏 ²)
1	1.5025	1.703×10^{-4}
2	1.5230	0.555×10^{-4}
3	1.5215	0.354×10^{-4}
4	1.5170	0.021×10^{-4}
5	1.5135	0.042×10^{-4}
6	1.5155	2.5×10^{-9}
7	1.5140	0.024×10^{-4}
8	1.5205	0.245×10^{-4}
9	1.5095	0.366×10^{-4}
10	1.5185	0.087×10^{-4}
$\sum_{i=1}^{10}$	15.1555	3.397×10^{-4}

$$1.06 \times 1.943 \times 10^{-3} \text{ 伏} \approx 0.0021 \text{ 伏,}$$

置信水平 90% 的误差限为

$$1.833 \times 1.943 \times 10^{-3} \text{ 伏} \approx 0.0036 \text{ 伏。}$$

由此可给出实验结果的报道: 实验测得电源电压值为

$$(1.5156 \pm 0.0021) \text{ 伏 } (N=10),$$

或

$$(1.5156 \pm 0.0036) \text{ 伏 (置信水平 90%, } N=10)。$$

这两种报道给出的是同一个实验结果, 只不过采取的置信水平不同而已。前者因取置信水平为 68.3%, 按惯例可以不注明, 但 $N=10$ 则是不可少的, 它反映了给出的报道不是按正态分布而是按自由度为 9 的 t 分布。

上述的对这个实验的数据处理是按未知标准误差的情况来对待的, 实际上用数字电压表来测量电压时其标准误差是已知的 (但不是正态分布)。我们对样本进行 χ^2 检验:

$$\chi^2 = \frac{\sum_{i=1}^{10} (v_i - \bar{v})^2}{\sigma^2} = \frac{3.397 \times 10^{-4}}{(0.00014)^2} \approx 1.6 \times 10^4。$$

它的期待值为 9, 因此 χ^2 值应在 9 的附近, 现在实际得到的 χ^2 值实在太大了, 需要对其原因进行认真的分析。数字电压表的误差分布是均匀分布而非正态分布, 这一点虽然会带来一些影响, 但不会那么大。实际上, χ^2 远大于其期待值的原因是对这个实验的标准误差给得偏小

了.上面提到数字电压表的标准误差是 0.00014 伏,这是指用数字电压表来测量某一确定的电压时的标准误差.问题在于现在测量的对象是一个特定的电源电压,而电源电压总有一定的不稳定性,表现为一定的随机起伏,它也是测量时统计误差的一个来源.在这个例子中,随机起伏所导致的统计误差远远大于仪器本身的标准误差,亦即这个实验的标准误差应该主要是由电源电压本身的统计起伏所决定.只考虑仪器的标准误差并认为就是实验的标准误差,这样作的后果是把实验的标准误差给得偏小了.

从这个例子还可以看到以下两点:

(1) 在仪器的标准误差为已知的情况下,进行样本 χ^2 检验还有助于判断实验结果的误差主要是由仪器的精密程度决定,还是主要由被测量的物理量本身的随机统计起伏或其它原因所决定.

(2) 在仪器的标准误差为已知的情况下,如果只进行一次或很少次数的测量,由于难于通过 χ^2 检验来考察被测物理量本身随机统计起伏的大小,就有可能用仪器的已知标准误差来标定实验结果的标准误差,从而大大夸大了实验结果的精确性.因此,在不允许作重复多次测量的实验中,在确定一次测量的误差时,需要特别注意避免发生这种问题.

三、间接测量与误差的传播

许多物理量是间接测量的,是利用直接测量的结果通过适当的公式推算出来的.在这时就需要将间接观测量的误差通过直接观测量的误差推算出来,这就是通常所说的误差的传播问题.下面分几方面进行讨论.

1. 如果间接观测量 y 是直接观测量 x_i ($i=1, \dots, N$) 的函数, $y = y(x_i)$, 将 $y(x_i)$ 在 $x_i = \langle x_i \rangle$ 附近作 Taylor 展开,并略去二次以上的高次项,得

$$y(x) = y(\langle x \rangle) + \sum_i \left(\frac{\partial y}{\partial x_i} \right)_{x=\langle x \rangle} (x_i - \langle x_i \rangle).$$

用这个近似表达式可以求出 y 的方差为

$$\begin{aligned} \sigma^2(y) &= \sum_i \sum_j \left(\frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j} \right)_{x=\langle x \rangle} \text{Cov}(x_i, x_j) \\ &= \sum_i \left(\frac{\partial y}{\partial x_i} \right)_{x=\langle x \rangle}^2 \sigma^2(x_i) + 2 \sum_{i < j} \cdot \\ &\quad \left(\frac{\partial y}{\partial x_i} \frac{\partial y}{\partial x_j} \right)_{x=\langle x \rangle} \text{Cov}(x_i, x_j). \end{aligned} \quad (3.9)$$

这就是误差传播的基本公式.在实际应用时,常常利用 \bar{x} 近似代替 $\langle x \rangle$, 利用实验测得的方差和协方差的估值代替方差与协方差.

这个公式实际上可以用下述比较容易记忆的规则简便地给出:

对函数 $y = y(x)$ 求微分,得

$$dy = \sum_i \frac{\partial y}{\partial x_i} dx_i,$$

将等式两边平方,然后作如下的代换,

$$\begin{aligned} (dy)^2 &\rightarrow \sigma^2(y), & (dx_i)^2 &\rightarrow \sigma^2(x_i), \\ dx_i dx_j &\rightarrow \text{Cov}(x_i, x_j). \end{aligned}$$

代换之后,再将其中出现的 $\frac{dy}{dx_i}$ 中的 x_i 一律用 \bar{x}_i 代入,将 $\sigma^2(x_i)$ 和 $\text{Cov}(x_i, x_j)$ 一律用实验所给出的估值代入,就得到 $\sigma^2(y)$ 的表达式 (3.9) 在实际应用中的形式.

如果各 x_i 是互相独立的,则 $\text{Cov}(x_i, x_j) = 0$, 上述规则就更简化,即

$$dy = \sum_i \frac{\partial y}{\partial x_i} dx_i$$

直接对应成

$$\sigma(y) = \sqrt{\sum_i \left(\frac{\partial y}{\partial x_i} \right)^2 \sigma^2(x_i)}, \quad (3.10)$$

亦即可以简单归结为:求微分后,求和号内平方,求和号外再开方,所有的微分改成对应的标准误差.

现在举一个例子.在高能物理学中如果观察到一个粒子,测出它的能量 E 和动量 P , 就可以利用

$$m^2 = \left(\frac{E}{c^2} \right)^2 - \left(\frac{P}{c} \right)^2$$

计算出这个粒子的质量 m . 要估算质量的标准误差,可对上式求微分,得

$$mdm = \frac{1}{c^4} E dE - \frac{1}{c^2} P dP.$$

如果在实验中能量和动量是独立进行测量, 则误差公式相应为

$$\sigma(m) = \frac{1}{mc^2} \sqrt{\left(\frac{1}{c^2} E\right)^2 \sigma^2(E) + P^2 \sigma^2(P)}.$$

这里 E 和 P 就用测得的值代入, m 则是利用上式由测得的 E 和 P 算出的, 这样就可以从测量 E 和 P 的标准误差推算出质量的标准误差.

如果直接观测量 $x_i (i = 1, \dots, N)$ 是互相独立的, 则从上面给出的误差传播公式可得到以下几个推论:

(1) $y = \sum_i x_i$ 的误差由下式给出:

$$\sigma(y) = \sqrt{\sum_i \sigma^2(x_i)}. \quad (3.11)$$

(2) $y = \prod_i x_i$ 的误差由下式给出:

$$\frac{\sigma(y)}{|y|} = \sqrt{\sum_i \left[\frac{\sigma(x_i)}{|x_i|}\right]^2}. \quad (3.12)$$

(3) $y = x^n$ 的误差由下式给出:

$$\frac{\sigma(y)}{|y|} = \frac{n}{|x|} \sigma(x). \quad (3.13)$$

根据(3.13)式, (3.12)式也就可以推广到包括乘和除在内的普遍情形. 在实际计算时, 这些公式中出现的期待值往往用样本平均值来代替.

要强调的是这些结果都是在各 x_i 值互相独立的假定下得出的. 如果不互相独立, 结果就不一样了. 例如对于 $y = x^2$, 按(3.13)式得到的是 $\frac{\sigma(y)}{|y|} = 2 \frac{\sigma(x)}{|x|}$, 这是正确的. 但如果把两个 x 看作是互相独立的物理量而采用(3.12)式进行计算, 则得到

$$\frac{\sigma(y)}{|y|} = \sqrt{2} \frac{\sigma(x)}{|x|}.$$

这表明, 由于把两个 x 错误地看作是互相独立的物理量, 造成的后果是把误差估计偏小了.

作为一个例子, 让我们看看关于电子质量的间接测量问题. 在核物理学中常把电子质量 m_e 乘光速 c 的平方用能量的单位 MeV 来度量. 在原子光谱学中常用到 Rydberg 常数 R_∞ , 可表示为

$$R_\infty = \frac{m_e c^2 \alpha^2}{4\pi \hbar c}.$$

这样 $m_e c^2$ 可表为

$$m_e c^2 = 4\pi(\hbar c) R_\infty \alpha^{-2},$$

其中 $\hbar = \frac{1}{2\pi} h$, h 为 Planck 常数, α 是精细结构常数.

$m_e c^2$ 是很难直接测准的, 但可以通过测准 $(\hbar c)$, R_∞ 和 α 来给出 $m_e c^2$ 准确的估值. 在这个式子中, 各直接测量的物理量是互相独立的, 可以用上面(3.12)式和(3.13)式直接给出误差计算公式

$$\frac{\sigma(m_e c^2)}{m_e c^2} = \sqrt{\left[\frac{\sigma(\hbar c)}{\hbar c}\right]^2 + \left[\frac{\sigma(R_\infty)}{R_\infty}\right]^2 + \left[2 \frac{\sigma(\alpha^{-1})}{\alpha^{-1}}\right]^2}.$$

在这里, 式中所有的期待值的符号都已略去.

三个测得较准的物理量是 $\hbar c$, R_∞ 和 α^{-1} , 其目前通用的实验值及由此给出的相对误差分别为

$$\hbar c = (1.9732858 \pm 0.0000051) \times 10^{-11} \text{MeV} \cdot \text{cm}^{-1},$$

$$\frac{\sigma(\hbar c)}{\hbar c} = 2.6 \times 10^{-6},$$

$$R_\infty = (109737.3177 \pm 0.0083) \text{cm}^{-1},$$

$$\frac{\sigma(R_\infty)}{R_\infty} = 0.075 \times 10^{-6},$$

$$\alpha^{-1} = (137.03604 \pm 0.00011),$$

$$\frac{\sigma(\alpha^{-1})}{\alpha^{-1}} = 0.82 \times 10^{-6}.$$

由此推算得 $m_e c^2$ 的估值为 0.5110034MeV , 相对误差为 $\frac{\sigma(m_e c^2)}{m_e c^2} = 2.8 \times 10^{-6}$, 由此得到实验给出间接测量 $m_e c^2$ 的结果为

$$m_e c^2 = (0.5110034 \pm 0.0000014) \text{MeV}.$$

这就是现在公认的通用的电子质量的实验值.

但是在 1973 年以后的实验中, 测量 R_∞ 和 α^{-1} 的精度都有了提高, 其值分别为

$$R_\infty = (109737.31521 \pm 0.00011) \text{cm}^{-1},$$

$$\frac{\sigma(R_\infty)}{R_\infty} = 0.001 \times 10^{-6},$$

$$\alpha^{-1} = 137.035963 \pm 0.000015,$$

$$\frac{\sigma(\alpha^{-1})}{\alpha^{-1}} = 0.11 \times 10^{-6}.$$

如果改用这些值来推算电子的质量, 结果为

$$m_e c^2 = (0.5110028 \pm 0.0000013) \text{MeV},$$

其精度比以前略有提高, 但现在公认的通用的值仍是前面给出的值.

2. 在实验中同时有几个直接观测量时, 如何判断它们是否是互相独立的? 如果事先无法知道它们是否是互相独立的, 则只有依据样本来估计它们的协方差, 并利用普遍的误差传播公式 (3.9) 来进行计算.

在前面已讲到, 当测量次数 N 较大时, 可以用样本平均值的标准偏差

$$S_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N(N-1)}}$$

作为 $\sigma(\bar{x})$ 的估计值, 亦即用 $S_{\bar{x}}^2$ 作为方差 $\sigma^2(\bar{x})$ 的估计值. 现在考虑两个直接观测量 x 和 y , 作了 N 次测量, 得到 x 和 y 的样本平均值分别为

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i, \quad \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i.$$

定义样本共差为

$$S_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}). \quad (3.14)$$

它的期待值为

$$\begin{aligned} \langle S_{xy} \rangle &= \frac{1}{N-1} \sum_{i=1}^N \langle (x_i - \bar{x})(y_i - \bar{y}) \rangle \\ &= \frac{N}{N-1} \langle (x_1 - \bar{x})(y_1 - \bar{y}) \rangle \\ &= \frac{1}{(N-1)N} \left\langle \left[(N-1)x_1 - \sum_{i=2}^N x_i \right] \cdot \right. \\ &\quad \left. \left[(N-1)y_1 - \sum_{i=2}^N y_i \right] \right\rangle. \end{aligned}$$

注意到 $\langle x_i y_i \rangle = \langle xy \rangle$, $\langle x_i y_j \rangle = \langle x \rangle \langle y \rangle$ (当 $i \neq j$ 时),

$$\langle S_{xy} \rangle = \langle xy \rangle - \langle x \rangle \langle y \rangle = \text{Cov}(x, y). \quad (3.15)$$

和方差的情形相对应, 样本平均值的标准偏差

为

$$S_{\bar{x}\bar{y}} = \frac{1}{N} S_{xy}, \quad (3.16)$$

其期待值为

$$\langle S_{\bar{x}\bar{y}} \rangle = \frac{1}{N} \text{Cov}(x, y) = \text{Cov}(\bar{x}, \bar{y}). \quad (3.17)$$

在 N 很大时, 可以用样本平均值的标准偏差作为标准误差的估计值, 同样地可以用样本平均值的共差 S_{xy} 作为协方差的估计值.

现在的问题是如果 N 不够大, 这种估计值与现实偏离的程度可能有多大. 为此需要计算 $S_{\bar{x}}^2$ 和 S_{xy} 的方差, 这标识它们在期待值附近摆动的幅度, 也就能反映在用它们代替期待值时实际可能偏离的大小. 若 x 和 y 服从正态分布, 并且样本容量 N 较大时, 有

$$\sigma(S_{\bar{x}}^2) = \sqrt{\frac{2}{N-1}} \frac{\sigma^2}{N}, \quad (3.18)$$

$$\sigma(S_{xy}) = \frac{1}{\sqrt{2}} \frac{\sigma}{N}, \quad (3.19)$$

值得注意的是 $S_{\bar{x}}^2$ 是作为 $\frac{\sigma^2}{N}$ 的估计值, 而 S_{xy} 是作为 $\frac{\sigma}{\sqrt{N}}$ 的估计值, 因此这些估计值的相对误差为

$$\frac{\sigma(S_{\bar{x}}^2)}{S_{\bar{x}}^2} = \sqrt{\frac{2}{N-1}}, \quad (3.20)$$

$$\frac{\sigma(S_{xy})}{S_{xy}} = \frac{1}{\sqrt{2N}}.$$

S_{xy} 的计算结果大体相同, 只不过繁杂一些. 从这个结果来看, 用样本平均值的标准偏差作为标准误差的估计值时, 其相对误差为 $\frac{1}{\sqrt{2N}}$. 这

要求测量次数愈多愈好. 例如 $N=8$ 时, 相对误差达 0.25, 在 $N=50$ 时, 相对误差为 0.10.

前面已讲到, 当 N 不是很大时, 考虑到 $S_{\bar{x}}$ 的随机性, 误差的报道应根据 t 分布的理论来进行.

如果 y 是间接观测量, 并且是几个直接观测量的函数, 当处理数据时, 若忽视了样本共差的估计而误把它们当作互相独立的量来处理, 则得到的 y 的标准误差的估计值就可能不正

确,也可能偏大,也可能偏小。

现在考察这种疏忽所带来的影响。若物理量 y 是通过两个直接观测量 x_1 和 x_2 的测量结果推算的,如果由于 x_1 的随机性带来 y 的误差的估计值为 σ_1 , x_2 带来的误差为 σ_2 , 在忽视了共差处理的情况下,给出 y 的标准误差的估计值为

$$\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}.$$

如果共差不为零,正确的估计值 σ 应在区间 $[|\sigma_1 - \sigma_2|, \sigma_1 + \sigma_2]$ 之内。换言之,忽视共差对标准误差估计值上带来的偏离大小决定于 σ_1 和 σ_2 中的小者,而标准误差估计值的大小则主要由 σ_1 和 σ_2 中的大者决定。这也说明,如果 σ_1 与 σ_2 两者相差悬殊,则忽视共差处理对误差估计的影响就很小。但在实验工作中处理数据时,不应忘记共差处理。特别是作共差处理时,并不需要作额外的测量,利用直接测量得到的样本就已经足够了。

3. 现在来考察上面给出的误差传播公式的应用范围。(3.9)式的基础是 Taylor 展开,并且只保留到一级微商项,这相当于在期待值附近把函数用线性函数来近似描写。显然,这样做的前提是要求误差实际上比较小,这相当于 Taylor 展开时离展开点不远,高次项可以略去。另外要看函数在展开点附近的行为:如果函数在展开点达极值或接近极值,则展开式中一次项为零或接近于零,这时二次项的贡献可能并不能忽略;如果函数在展开点附近变化迅速,则高次项的贡献应较大,从而也不应轻易加以忽略。总之,有上述各种情况出现时,特别是直接测量结果中有些量的相对误差较大时,就要特别小心,仔细考察上面给出的公式是否能用。

在不能用(3.9)式到(3.13)式时怎么办?原则上是通过 y 与直接观测量 x_i 之间的函数关系和各 x_i 的概率密度函数推算出 y 的概率密度函数,然后根据要求的置信水平给出关于 y 的实验结果的报道。举一个最简单的例子,如果 $y = x^{-3}$, x 的直接测量结果是 $x = 1.00 \pm 0.20$, 这表明 x 的真值有 68.3% 的概率在 0.80 到 1.20 之间。如果要在同样的置信水平下来报

道 y 的测量值时,可以把 $x=0.80, 1.00$ 和 1.20 时相应的 y 值求出,它们为 $y = 1.95, 1.00$ 和 0.58 , 并由此给出 y 的测量值的报道为 $y = 1.00 \pm 0.20$ 。注意这时误差是不对称的,这个式子的含义是 y 的期待值的估计值为 1.00, y 的真值在 1.00 到 1.95 内和在 0.58 到 1.00 内的概率分别都是 0.5×0.683 。因此尽管误差是不对称的,但置信水平仍是 68.3%, 并且在估计值两侧的区间内置信水平是相同的。

当 y 是多个变量的函数时,如果不能采用(3.9)式来近似估算 y 的误差,这时要得到 y 的误差的处理过程就要复杂得多,主要是要设法求 y 的概率密度函数,在第二讲中我们给出了求 y 的概率密度函数的普遍公式(2.29)。

四、本底估算和本底扣除

在近代物理实验(特别是一些精确实验)中遇到的一个很重要的问题是本底的处理。本底可以概括为:在实验中为研究某物理现象而测量能反映该物理现象的某些物理量时,本实验所研究的物理现象以外的各种因素对所测物理量给出的一切贡献,并且当这部分贡献与所要研究物理现象的贡献混在一起被测量时,则这部分贡献称为本底。

本底可以来自其它物理现象,也可能来自实验中所用仪器设备性能上的特点。本底本身与所要研究的物理现象无关,但本底却又和有用的实验结果混在一起。如果实验中不注意把本底和有用的结果区分开来,就有可能从实验中得出错误的判断。

本底的表现形式是多种多样的,处理的方法也有所不同,下面作几点讨论:

1. 要把本底和可扣除的系统误差区分开来,在实验设计时设法把可以消除的系统误差消除掉,并且在处理数据时把没有消除但可以扣除的系统误差尽量予以扣除,使无法扣除的都只是本底。实际上,本底的表现形式往往与难以扣除的系统误差相似。但是一般的本底之所以难以扣除,其原因往往(但并不都是)在于

物理现象本身,而不只是由测量手段带来的.在把可以扣除的系统误差都扣除掉之后,就可以集中处理本底问题了.

2. 在实验中需要对本底进行估算,然后在实验结果中将本底扣除以得到所需要的结果.最常用的方法是另外再做一部分实验单独测量本底.也就是说,如果要观测的物理现象 A 会对物理量 x 有贡献,实验是通过测量 x 而研究 A ,则可以在没有 A 现象出现时测量物理量 x ,测得的结果作为本底.这样测量本底之后,再把有 A 现象出现时测量 x 的结果减去本底,就得到直接由于现象 A 而贡献的 x 的部分,这就是实验所需要得到的结果.

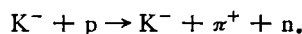
这个方法最常用,也已经标准化了.在用此方法处理时,要注意以下两点:

(1) 测量本底时实验条件中除了去掉现象 A 外,其它条件都应不变,以保证测得的本底可靠;

(2) 本底的数据处理以及从测量结果中扣除本底都应按照误差理论的要求进行.

3. 有时由于实验比较复杂,或者由于所研究的物理现象的性质,使得难以单独测量本底,这就需要从实验结果中设法直接扣除本底.在这种情况下,一种常用的方法是研究伴随现象 A 所贡献的物理量 x 还可能有哪些特征是其它原因对 x 的贡献所没有的,利用这个特征来把本底部分确定下来,并在结果中予以扣除.

举一个典型例子.在六十年代初期,发现了大量的新粒子.这些新粒子的特点是极不稳定,大多数的平均寿命为 10^{-22} 秒的量级.一般说来,在它们产生之后还未走出原子的范围就很快衰变掉了.实验上发现这种短寿命粒子的手段是通过对这种粒子衰变产物的测量推测其存在.例如,在用 K^- 介子撞击质子 p 的实验中,把反应后生成中子 n 加 K^- 介子再加 π^+ 介子的事例挑选出来,即反应为



如果 K^- 和 π^+ 是某一种新粒子衰变成的,则可以利用 K^- 介子的动量 P_K 和能量 E_K , 以及 π^+ 介子的动量 P_π 和能量 E_π 推算出这种新粒子的

物理

质量 m 满足

$$m^2 = \left(\frac{E_K + E_\pi}{c^2} \right)^2 - \left(\frac{P_K + P_\pi}{c} \right)^2.$$

把所有的事例中对应的 m^2 都算出来,然后按下面的方法统计作图.以 m^2 为横轴,以单位 m^2 间隔所观察到的事例数为纵轴,得到的图形可以是图 1 所示的三种之一.第一种是所有事例中 $K + \pi$ 都是由这种新粒子衰变成的,如图 1 (a) 所示;第二种是没有事例是由新粒子衰变而产生 $K + \pi$ 的,其图形如图 1(b) 所示;第三种是有一部分事例中的 $K + \pi$ 是由新粒子衰变而产生的,其图形如图 1(c) 所示.

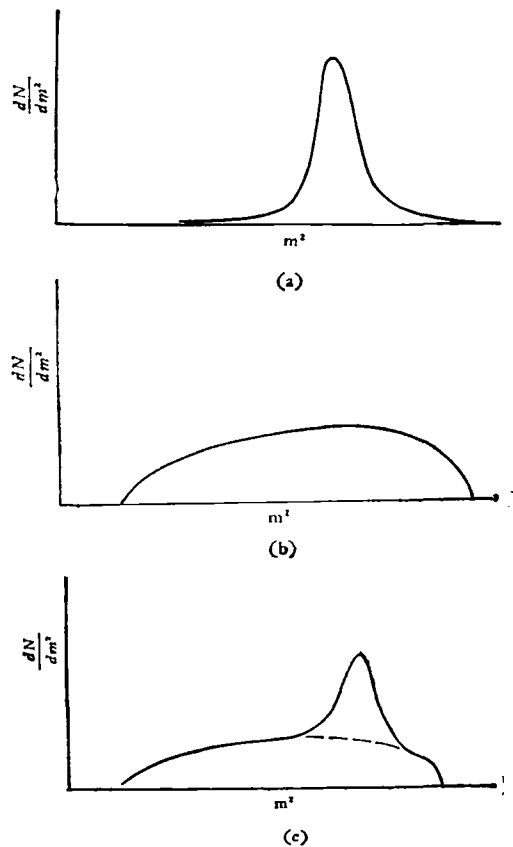


图 1

实际观察到的结果是属于上述第三种情形.在这个实验中,要研究的现象 A 是:“是否存在新粒子,其衰变产物为 $K + \pi$ ”.观察的物理

量 x 是一个质量值,它的含义是:“如果 $K + \pi$ 是由一个新粒子衰变而来的 x 就是这个新粒子的质量”,处理的方法是给出事例按质量(或质量平方)的分布图。采取这种处理方法的原因,是由于新粒子的存在将在质量分布图上表现为一个“峰”,而本底则表现为平滑曲线。利用这个特征,在实际得到图 1(c) 的结果时,可以把本底加以区分和扣除,并且把实验结果中本底所占的比例也能确定下来。

在这种方法中,找出物理现象 A 所贡献的 x 具有的特征是很关键的一步,是分析和处理数据的基础。在上面这个例子中,这个特征在质量分布图上表现为“峰”。但是在作这种分析时,还应注意避免假象。在上面这个例子中,还可以对 $\pi + n$ 也做这样的分析。如果没有衰变为 $\pi + n$ 的新粒子存在,在 $\pi + n$ 系相应的质量分布图中应表现为平滑曲线。但是由于在 $K + \pi$ 系统表现出峰来,在 $\pi + n$ 系统中也会看到一个“峰”。这个峰称为衰变为 $K + \pi$ 的新粒子的“影子”,因为它并不反映真有一个可以衰变为 $\pi + n$ 的新粒子存在。

影子的产生可以用 $K + \pi$ 系和 $\pi + n$ 系的质量分布图来说明。如图 2 所示,每一个事例的 $K + \pi$ 系和 $\pi + n$ 系的不变质量 $m(K + \pi)$ 和 $m(\pi + n)$ 都可以算出来,从而在纵轴为 $m(\pi + n)$ 横轴为 $m(K + \pi)$ 的图上表现为一个点。图上的圈表示运动学允许的边界,因此实验观测到的事例的点一定落于圈内。如果没

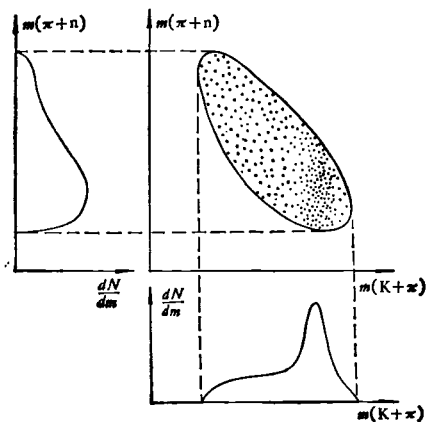


图 2

有新粒子存在,观察大量事例后,反映这些事例的点应在运动学允许的圈内,大体上表现为均匀分布。新粒子的存在则应表现为一些点密集在一个水平的或垂直的带上。如图 2 所示,有许多事例点密集在一个垂直的带上,这反映了在这个反应中有一部分事例出现了一种可以衰变为 $K + \pi$ 的新粒子。这个新粒子的存在,表现为图 2 给出的 $m(K + \pi)$ 分布图中出现的一个峰,但同时 $m(\pi + n)$ 分布图中也表现一个“峰”状隆起,这就是影子,如果不注意的话,就可能会误认为在 $\pi + n$ 系也反映出一种可以衰变为 $\pi + n$ 的新粒子存在。

由此可见在用这种方法处理时,在找到实验中可以区分本底的特征并根据这个特征来设计处理方案时,必须同时注意分析可能出现的影子或其它假象,以避免作出错误的判断。

4. 如果需要从实验结果中设法扣除本底,同时又找不到区别本底的特征,则需要对本底进行严格的理论估算。这是严格细致的理论分析工作,其要点是:

- (1) 根据实验处理的问题分析,列举出所有可能对本底有贡献的物理机制;
- (2) 逐个计算每个物理机制在本实验中对本底的贡献;
- (3) 把所有的本底相加,得到总的本底估算;
- (4) 如果本底的计算结果具有概率统计性质,则在计算过程中,包括各个机制贡献的本底相加时都必须用概率统计描述,以求准确地反映实际。

注意在作这样的估算时,不要简单地用“四舍五入”之类的近似。因为在计算过程中人为地引入的误差,常常会导致不应有的错误。特别是如果本底是不连续的量(例如上例中的事例数等)时,不要因为本底估算出来的值不是整数而轻易地四舍五入。一般说来,在估算的中间过程中应尽可能不作近似,在得到最后本底的总的估算值后再作可以允许的近似才是恰当的做法。