第五章 统计量的分布和测量误差理论

张松
核科学与技术系|现代物理研究所

第五节 不等精度观测的误差处理

所谓等精度测量就是指各个观测值相应的分布有相同的方差。前面所介绍的测量误差的处理和表示方法，就是在等精度测量的情况下进行的。一般认为，在相同的条件下，用同样的仪器和同样方法测得的数据是等精度的。在不同条件下用不同精度的仪器或不同方法所测得的数据是不等精度的

- 权的概念和加权均值
- 单位权方差的估计
- 数据协调性的检验

权的概念和加权均值

- 处理不等精度的观测数据时，对精度较高的数据应该给于较大的信赖和重视。为了反映对各个数据的不同信赖或重视程度，需要引进权的概念，赋给各个数据一个“权”，“权”大的数据确定最后的结果时起较大的作用，方法?

- **最大似然法**。设一袋中装有两种颜色的球，白球和黑球。已知黑球比白球多，试估计二者数量之比。这个问题也可以看做是要估计随机摸得白球的概率p'。自然摸得黑球的概率是$1-p'$。为此我们做放回抽样，即每次摸球后都放回袋中，摇匀后再摸第二次。设总共摸球一百次，得到一个样本$x_1, x_2, \cdots, x_{100}$，其中9次摸到白球，91次摸到黑球。由于每次摸球都是独立进行的，每一个x_i都可视为一独立的随机变量，故实现这样一个特定样本的联合概率密度为$p(x_1, x_2, \cdots, x_{100}; p') = p'^9(1-p')^{91}$，**这个样本的联合概率密度又叫做样本的似然函数**

- 可以这样设想，如果一次试验就出现了这个特定的样本，就有理由认为这个样本出现的可能性最大。因此，p'的估计值应该使似然函数值达到最大。应用求函数极值的微分法，很容易求得$p' = 0.09$

- 这个例子采用的方法就是所谓的最大似然法

权的概念和加权均值（续）

- 设有N个不等精度的互相独立的观测值$(x_1, x_2, \cdots, x_N)$，它们的方差分别用$\sigma_1^2, \sigma_2^2, \cdots, \sigma_N^2$来表示。采用最大似然法来估计待测量的真值$\mu$，将样本的似然函数表示为
$L(\underline{x}\backslash\mu) = L(x_1, x_2, \cdots, x_N\backslash\mu)$

- 样本的**似然函数**就是样本$(x_1, x_2, \cdots, x_N)$的**联合概率密度函数**，似然函数值就是N次测量得到的样本$(x_1, x_2, \cdots, x_N)$的联合概率密度值，$p(x_1, x_2, \cdots, x_N) = p(x_1)p(x_2)\cdots p(x_N)$，因此，**似然函数应为各测量值概率密度之积**

权的概念和加权均值（续）

- 根据中心极限定理，观测值x_i近似服从正态分布，设$(x_1, x_2, \cdots, x_N)$是对同一固定量μ的N个具有不同精度的观测值，x_i应服从期待值为μ，方差为σ_i^2的正态分布，即

$$p(x_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma_i}\right)^2\right]$$

则，样本$\underline{x} = (x_1, x_2, \cdots, x_N)$的似然函数应为

$$\begin{aligned} L(\underline{x}\backslash\mu) &= p(x_1, x_2, \cdots, x_N) \\ &= \prod_{i=1}^{N} p(x_i; \mu, \sigma_i^2) \\ &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma_i}\right)^2\right] \end{aligned} \qquad (5.42)$$

- 对不同的μ值，上式代表的概率是不同的，我们目的在于**寻求这样一个μ的估计值，使得上式取最大值**，这就是前面提到的最大似然法，为此必须解如下似然方程，注意：上式右边含指数，为求解方便，常用$\ln L(\underline{x}\backslash\mu)$代替$L(\underline{x}\backslash\mu)$作为似然函数

权的概念和加权均值（续）

- 似然方程$\dfrac{d\ln L(\underline{x}\backslash\mu)}{d\mu}=0 \quad (5.43)$

  不难求得$\ln L(\underline{x}\backslash\mu)$取极值的条件为$\sum_{i=1}^{N}\dfrac{x_i-\mu}{\sigma_i^2}=0 \quad (5.44)$

  由上式可解得μ的估计值

$$\hat{\mu}=\frac{\sum_{i=1}^{N}(x_i/\sigma_i^2)}{\sum_{i=1}^{N}1/\sigma_i^2}=\frac{\sum_{i=1}^{N}\omega_i x_i}{\sum_{i=1}^{N}\omega_i}\equiv\bar{x} \quad (5.45)$$

- 式中$\omega_i=1/\sigma_i$，称作观测值x_i的权，$\hat{\mu}$称作加权平均值，仍用$\bar{x}$表示。上式表明，**权越大（即方差越小）的观测值x_i在计算加权平均值$\hat{\mu}$时起的作用越大，达到了预期目的**。对于等精度测量，所有观测值的方差相同，因而权也相同，是上式的特殊情况

- 对加权平均值公式分子分母同乘以一个任意的正数σ^2，其结果不变，也就是说可令权

$$\omega_i=\frac{\sigma^2}{\sigma_i^2},\ i=1,2,\cdots,N \quad (5.46)$$

- 因此，权只具有相对意义，如果把σ^2取为某一观测值的方差σ_i^2，该观测值的权就等于1。可见σ^2相当于权为1的观测值的方差，称为**单位权方差**。若取单位权方差$\sigma^2=1$，则

$$\omega_i=\frac{1}{\sigma_i^2} \quad (5.46')$$

权的概念和加权均值（续）

- 加权平均值是观测值的一个线性组合，则其期望值

$$\langle \bar{x} \rangle = \left\langle \frac{\sum_{i=1}^{N} \omega_i x_i}{\sum_{i=1}^{N} \omega_i} \right\rangle = \frac{\sum_{i=1}^{N} \omega_i \langle x_i \rangle}{\sum_{i=1}^{N} \omega_i} = \mu \quad (5.47)$$

- 因x_i间相互独立，则加权平均值的方差

$$\sigma^2(\bar{x}) = \sigma^2 \left(\frac{\sum_{i=1}^{N} \omega_i x_i}{\sum_{i=1}^{N} \omega_i} \right) = \frac{1}{(\sum_{i=1}^{N} \omega_i)^2} \sigma^2 \left(\sum_{i=1}^{N} \omega_i x_i \right)$$

$$= \frac{1}{\left(\sum_{i=1}^{N} \omega_i\right)^2} \sum_{i=1}^{N} \omega_i^2 \sigma_i^2 = \frac{1}{\left(\sum_{i=1}^{N} \omega_i\right)^2} \sum_{i=1}^{N} \omega_i \sigma^2$$

$$= \frac{\sigma^2}{\sum_{i=1}^{N} \omega_i} = \frac{1}{\sqrt{\sum_{i=1}^{N} 1/\sigma_i^2}} \quad (5.48)$$

没有平方

权的概念和加权均值（续）

- 加权平均值$\bar{x}$的标准误差为

$$\sigma_{\bar{x}} = \sqrt{\sigma^2(\bar{x})} = \frac{\sigma}{\sqrt{\sum_{i=1}^{N} \omega_i}} = \frac{1}{\sqrt{\sum_{i=1}^{N} 1/\sigma_i^2}} \quad (5.49)$$

- 加权平均值$\bar{x}$的权是$\omega_{\bar{x}} = \dfrac{\sigma^2}{\sigma^2(x)} = \displaystyle\sum_{i=1}^{N} \omega_i \quad (5.50)$

- 由于观测值x_i服从正态分布，$x_i \sim n(x; \mu, \sigma^2)$，而加权平均值$\bar{x}$是观测值$x_i$的线性函数，则$\bar{x}$也应服从正态分布$\bar{x} \sim n(\bar{x}; \mu, \sigma^2(\bar{x}))$，则在加权平均值$\bar{x}$及其方差$\sigma^2(\bar{x})$计算后，可如下给出**观测报道**：

$$\mu = \bar{x} \pm \sigma(\bar{x}) \quad (5.51)$$

该报道是具有正态分布的概率意义

例题

- 对一放射源分别用小立体角法、$4\pi\beta$法、$\beta-\gamma$符合法测量它的放射性活度，其测得值分别为
  $9.32\times10^7 \quad 9.06\times10^7 \quad 8.99\times10^7\ (Bq)$

  若上述各结果的相对误差分别为$5.0\,\%$、$1.0\,\%$、$0.5\,\%$，试报道放射源的活度

- 设放射源的活度为A，则各观测值的方差为$(5A\%)^2 \quad (A\%)^2 \quad (0.5A\%)^2$

  取单位权方差$\sigma^2=(5A\%)^2$，可计算各观测值相应的权$\omega_i=\sigma^2/\sigma_i^2$为$1 \quad , \quad 25 \quad , \quad 100$

  放射性活度的加权平均值

  $$\bar{A}=\frac{\sum_{i=1}^{N}\omega_i A_i}{\sum_{i=1}^{N}\omega_i}=\frac{(1\times9.32+25\times9.06+100\times8.99)\times10^7}{1+25+100}=9.0\times10^7(Bq)$$

  $\bar{A}$的标准误差$\sigma(\bar{A})=\sqrt{\frac{\sigma^2}{\sum_{i=1}^{3}\omega_i}}=\sqrt{\frac{(5\times9.0\times10^5)^2}{1+25+100}}=4\times10^5\ (Bq)$

  最后测量结果报道为$A=\bar{A}\pm\sigma(\bar{A})=(9.00\pm0.04)\times10^7\ (Bq) \quad (68.3\%)$

单位权方差的估计

- 实际测量中，**各测量值的方差σ_i^2往往并不知道**，这样如何估计加权平均值的方差?

- 设对某物理量获得了N个不等精度的独立观测值$x_1, x_2, \cdots, x_N$，权分别是$\omega_1, \omega_2, \cdots, \omega_N$。现在求各观测值及平均值的方差。

- 通常先求出单位权方差σ^2，然后由上述公式求出各观测值及平均值的方差。则**单位权方差的估计是要解决的问题**。

- 最大似然法估计单位权方差，把$\sigma_i^2 = \sigma^2/\omega_i$带入似然函数得到
$$L(x_1, x_2, \cdots, x_N; \mu, \sigma^2)\frac{(\omega_1\omega_2\cdots\omega_N)^2}{(2\pi)^{(}N/2)\sigma^N} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\omega_i(x_i-\mu)^2\right]$$

- $\ln L$对σ^2的导数求极限，可确定使得似然函数最大的σ^2值，
$$\frac{d\ln L}{d\sigma^2} = \frac{-N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{N}\omega_i(x_i-\mu)^2 = 0$$

- 上式中的期待值μ可用加权平均值$\bar{x}$代入，这样解出的σ^2先用S^2代替（**估计！**），
$$S^2 = \frac{1}{N}\sum_{i=1}^{N}\omega_i(x_i-\bar{x})^2$$

单位权方差的估计（续）

- 由于$\bar{x}$系由样本求得，是统计量，必须检验S^2是否σ^2的无偏估计，

$$\left\langle S^2 \right\rangle = \left\langle \frac{1}{N}\sum_{i=1}^{N}\omega_i(x_i-\bar{x})^2 \right\rangle = \frac{1}{N}\left\langle \sum_{i=1}^{N}\omega_i\left[(x_i-\mu)-(\bar{x}-\mu)\right]^2 \right\rangle$$

$$= \frac{1}{N}\left\langle \left[\sum_{i=1}^{N}\omega_i(x_i-\mu)^2 - 2\sum_{i=1}^{2}\omega_i(x_i-\mu)(\bar{x}-\mu)\sum_{i=1}^{N}\omega_i(\bar{x}-\mu)^2\right] \right\rangle$$

$$= \frac{1}{N}\left[\sum_{i=1}^{N}\omega_i\left\langle (x_i-\mu)^2\right\rangle - 2\sum_{i=1}^{N}\omega_i\left\langle (x_i-\mu)(\bar{x}-\mu)\right\rangle + \sum_{i=1}^{N}\omega_i\left\langle (\bar{x}-\mu)^2\right\rangle\right]$$

因观测值x_i间相互独立，有 $\begin{array}{l}\left\langle (x_i-\mu)(x_j-\mu)\right\rangle = 0 \quad i\neq j \\ \left\langle (x_i-\mu)(x_j-\mu)\right\rangle = \sigma_i^2 \quad i=j\end{array}$，故

$$\left\langle (x_i-\mu)(\bar{x}-\mu)\right\rangle = \left\langle (x_i-\mu)\left(\frac{\sum_{j=1}^{N}\omega_j x_j}{\sum_{j=1}^{N}\omega_j}-\mu\right)\right\rangle$$

$$= \left\langle (x_i-\mu)\frac{\sum_{j=1}^{N}\omega_j(x_j-\mu)}{\sum_{j=1}^{N}\omega_j}\right\rangle = \frac{\sum_{j=1}^{N}\left[\omega_j\left\langle (x_i-\mu)(x_j-\mu)\right\rangle\right]}{\sum_{i=1}^{N}\omega_i}$$

$$= \frac{1}{\sum_{i=1}^{N}\omega_i}\omega_i\sigma_i^2 = \frac{\sigma^2}{\sum_{i=1}^{N}\omega_i} \qquad \text{(I)}$$

单位权方差的估计（续）

由(5.48)可得，$\left\langle (\bar{x}-\mu)^2 \right\rangle = \sigma^2(\bar{x}) = \dfrac{\sigma^2}{\sum_{i=1}^{N}\omega_i}$ (II)

另有$\left\langle (x_i-\mu)^2 \right\rangle = \sigma_i^2 = \dfrac{\sigma^2}{\omega_i}$ (III)

- 将上述推导整理代入$\left\langle S^2 \right\rangle$，得到

$$\left\langle S^2 \right\rangle = \frac{1}{N}\left[\sum_{i=1}^{N}\omega_i\left(\frac{\sigma^2}{\omega_i}\right) - 2\sum_{i=1}^{N}\omega_i\frac{\sigma^2}{\sum_{i=1}^{N}\omega_i} + \sum_{i=1}^{N}\omega_i\frac{\sigma^2}{\sum_{i=1}^{N}\omega_i}\right]$$
$$= \frac{1}{N}(N\sigma^2 - 2\sigma^2 + \sigma^2) = \frac{N-1}{N}\sigma^2$$

- 上式结果表明S^2是单位权方差σ^2的有偏估计，需作线性校正。因此**单位权方差σ^2的估计值应为**

$$\hat{\sigma}^2 = \frac{N}{N-1}S^2 = \frac{1}{N-1}\sum_{i=1}^{N}\omega_i(x_i-\bar{x})^2 \quad (5.52)$$

- 显然$\left\langle \hat{\sigma}^2 \right\rangle = \left\langle \dfrac{N}{N-1}S^2 \right\rangle = \sigma^2$，即$\hat{\sigma}^2$是单位权方差$\sigma^2$的无偏估计

单位权方差的估计（续）

- 利用(5.46)式可得到任意一观测值的方差估计

$$\sigma_i^2 = \frac{\hat{\sigma}^2}{\omega_i} = \frac{1}{(N-1)\omega_i}\sum_{i=1}^{N}\omega_i(x_i - \bar{x})^2 \quad (5.53)$$

- 加权平均值的样本方差

$$S_{\bar{x}}^2 = \frac{\hat{\sigma}^2}{\sum_{i=1}^{N}\omega_i} = \frac{1}{(N-1)\sum_{i=1}^{N}\omega_i}\sum_{i=1}^{N}\omega_i(x_i - \bar{x})^2 \quad (5.54)$$

- 加权平均值的标准偏差

$$S_{\bar{x}} = \sqrt{S_{\bar{x}}^2} = \sqrt{\frac{1}{(N-1)\sum_{i=1}^{N}\omega_i}\sum_{i=1}^{N}\omega_i(x_i - \bar{x})^2} \quad (5.55)$$

单位权方差的估计（续）

- 如果样本容量N较小(< 10)，测量结果应该按t分布报道
  $\mu = \bar{x} \pm t_\xi S_{\bar{x}}$ （置信水平ξ）

- 如果样本容量较大，测量结果可按正态分布报道
  $\mu = \bar{x} \pm u_\xi S_{\bar{x}}$（置信水平$\xi$）

- 如果σ_i^2已经知道，由于加权平均值是观测值的线性组合，根据定理2.2，定理2.3，加权平均值$\bar{x}$也应服从正态分布，其结果按(5.51)式作正态报道

数据协调性的检验

- 不等精度的观测数据常来自不同方法，或由不同实验者得到的测量结果。把他们放在一起整理分析及综合结果时常需要考虑所给出的数据和误差是否合理及测量中有无系统误差等问题，这就需要进行数据协调性检验。介绍两种方法

- **检验数据样本χ^2量$\sum_{i=1}^{N}\frac{(x_i-\bar{x})^2}{\sigma_i^2}$**。根据定理2.10，我们已经知道样本的$\chi^2$量服从自由度为$(N-1)$的$\chi^2$分布。因此$\chi^2$量的值应在$(N-1)$附近摆动。这说明所观测的诸数据$x_i$之间应该有一种联系。如果某些观测值$x_i$远远偏离真值，或某些观测值的误差估计太小，都可能使算出的样本χ^2值显著的偏离$(N-1)$。χ^2值的合理取值范围可参见“假设检验”相关章节。一般情况下，$\chi^2/(N-1)$在$0.5\sim2.0$之间即可认为是很不错的结果

数据协调性的检验（续）

- **比较单位权方差σ^2及其估计值$\hat{\sigma}^2$**。前面已经介绍过单位权方差σ^2的两种计算方法。一种是按(5.46)式有 $\sigma^2 = \omega_i \sigma_i^2 \quad (i = 1,2,\cdots,N)$。另一种按(5.52)式有 $\sigma^2 = \hat{\sigma}^2 = \dfrac{1}{N-1}\sum_{i=1}^{N}\omega_i(x_i - \bar{x})^2$。比较这两种方法得到的单位权方差是否一致，实质上也是检验样本χ^2量是否服从自由度为$(N-1)$的χ^2分布，因此可以用来对数据的协调性进行检验。一个经验性的具体方案是把$\hat{\sigma}/\sigma = 1 + u$，若满足$|u| \leqslant \dfrac{4}{3\sqrt{N-1}}$，则认为各数据间是协调的

例题

- 对30keV中子所测量的^{238}U(n,γ)俘获反应截面x_i(mb)及误差如下表所示。计算加权平均值及其标准误差，并检验数据的协调性。

i	x_i	σ_i	ω_i	$\omega_i x_i$	$\|x_i-\bar{x}\|$	$(x_i-\bar{x})^2$	$\omega_i(x_i-\bar{x})^2$
1	550	50	1.00	550.00	83.22	6925.98	6925.98
2	473	50	1.00	473.00	6.22	38.72	38.72
3	428	50	1.00	428.00	38.78	1503.69	1503.69
4	446	36	1.93	860.34	20.78	431.70	832.76
5	479	14	12.76	6109.69	12.22	149.39	1905.48
6	418	19	6.93	2894.74	48.78	2379.24	16476.76
7	469	24	4.34	2035.59	2.22	4.94	21.44
8	528	30	2.78	1466.67	61.22	3748.19	10411.65
9	465	23	4.73	2197.54	1.78	3.16	14.93
合计			36.45	17015.57			38131.42

例题（续）

解：取单位权方差$\sigma^2 = 50^2$，利用$\omega_i = \dfrac{\sigma^2}{\sigma_i^2}$

计算出各观测值相应的权ω_i，结果列在表中。
计算加权平均值

$$\bar{x} = \frac{\sum_{i=1}^{9} \omega_i x_i}{\sum_{i=1}^{9} \omega_i} = \frac{17015.57}{36.45} = 466.78$$

按(5.49)式计算加权平均值的标准误差

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{\sum_{i=1}^{9} \omega_i}} = \frac{50}{36.45} = 8.28$$

按(5.55)式计算加权平均值的样本偏差

$$S_{\bar{x}} = \hat{\sigma}_{\bar{x}} = \sqrt{\frac{1}{(N-1)\sum_{i=1}^{9}\omega_i}\sum_{i=1}^{9}\omega_i(x_i-\bar{x})^2} = \sqrt{\frac{38131.42}{(9-1)\times 36.45}} = 11.43$$

i	x_i	σ_i	ω_i	$\omega_i x_i$	$\lvert x_i - \bar{x}\rvert$	$(x_i - \bar{x})^2$	$\omega_i(x_i - \bar{x})^2$
1	550	50	1.00	550.00	83.22	6925.98	6925.98
2	473	50	1.00	473.00	6.22	38.72	38.72
3	428	50	1.00	428.00	38.78	1503.69	1503.69
4	446	36	1.93	860.34	20.78	431.70	832.76
5	479	14	12.76	6109.69	12.22	149.39	1905.48
6	418	19	6.93	2894.74	48.78	2379.24	16476.76
7	469	24	4.34	2035.59	2.22	4.94	21.44
8	528	30	2.78	1466.67	61.22	3748.19	10411.65
9	465	23	4.73	2197.54	1.78	3.16	14.93
合计			36.45	17015.57			38131.42

比较$\sigma_{\bar{x}}$和$S_{\bar{x}}$，取较大者$S_{\bar{x}}$，结果可报道为

$$\bar{x} \pm S_{\bar{x}} = 466.78 \pm 11.43 \text{ (mb)}$$

按两种不同方法计算单位权方差。这里
$\sigma = 50$

$$\hat{\sigma} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{9}\omega_i(x_i-\bar{x})^2} = \sqrt{\frac{38131.42}{9-1}} = 69.04$$

，$u = \dfrac{\hat{\sigma}}{\sigma} - 1 = 69.04/50 - 1 = 0.38$

由于

$$|u| = 0.38 < \frac{4}{3\sqrt{N-1}} = \frac{4}{3\sqrt{8}} = 0.47$$

，认为数据是协调合理的

作业

- 两次测得某物体的长度分别为 $1.47 \pm 0.02(\text{cm})$, $1.53 \pm 0.06(\text{cm})$，写出测量的最后结果和误差

- 设对某物理量所观测的数据和赋予的权为，$x_i = 0.507, 0.438, 0.381, 0.371, 0.350, 0.402,$ $\omega_i = 8,5,2,8,13,20$，报道该物理量的最后测量结果及其误差

- 测得的一组计数是1010,1018,1002,950,1060，试检验这组数据是否正常?