



误差理论与实验的数学处理讲座

第四讲 参数估计方法中的一些问题

高 崇 寿

(北京大学物理系)

一、多个实验结果的综合

先看一个最常用的参数估计问题的实例。在物理学的研究中,对于一个重要的物理量,常常要用多组实验进行测量。各组实验分别根据各自的结果给出报道,然后综合各组实验的结果,给出对该物理量最准确的报道。在这样作时,由于各个实验的精确度并不相同,因此各组实验的结果都应考虑进去,但又不应该简单地平等对待,而应根据各个实验的精确度不同而有所区别。这样的方法应该建立在概率论的基础上。

考虑一个物理量 x , 共有 N 个实验结果, 其观察值 x_i 服从正态分布, 实验结果的报道为

$$x = x_i \pm \sigma_i \quad (i = 1, \dots, N), \quad (4.1)$$

即所有的 x_i 的期待值都是同一个值 x (即物理量的真值), 其标准误差 σ_i 则随实验的不同而异。(4.1)式表明, 第 i 个实验给出 x 的估计值为 x_i , x 的真值处于 $x_i - \sigma_i$ 到 $x_i + \sigma_i$ 之间的概率是 68.3%。现在要解决的问题是如何在统一考虑这 N 组实验的基础上, 确定 x 的估计值及相应的标准误差。

现在首先看 $N = 1$ 的情形。 x_1 满足的是期待值为 x , 标准误差为 σ_1 的正态分布, 其概率密度函数为

$$p(x_1) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{(x_1 - x)^2}{2\sigma_1^2}\right].$$

我们说“ x 的估计值为 x_1 ”时, 可以这样理解: x_1 和 σ_1 是实验已定出来的值, x 是未知的, x 的

物理

估计值应使实验定出的 x_1 概率最大。换言之, 把 $p(x_1)$ 看作是 x 的函数, x 的估计值 \hat{x} 应该使 $p(x_1)$ 取极大值。按照这个要求, 由

$$\frac{\partial p(x_1)}{\partial x} = 0,$$

可立即定出 x 的估计值 $\hat{x} = x_1$, 该估计值的标准误差为

$$\sigma(\hat{x}) = \sigma(x_1) = \sigma_1.$$

这就说明了 $x = x_1 \pm \sigma_1$ 报道的来源。

下面把上述作法推广到 N 组实验的情形。

x_i 的概率密度函数为

$$p(x_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left[-\frac{(x_i - x)^2}{2\sigma_i^2}\right]. \quad (4.2)$$

x_1, \dots, x_N 的联合概率密度函数为

$$p(x_1, \dots, x_N) = \prod_{i=1}^N p(x_i). \quad (4.3)$$

x 的估计值 \hat{x} 应该使 $p(x_1, \dots, x_N)$ 作为 x 的函数取极大值。求微商得极大值条件为

$$\sum_{i=1}^N \frac{x_i - \hat{x}}{\sigma_i^2} = 0,$$

亦即 x 的估计值 \hat{x} 应为

$$\hat{x} = \frac{\sum_{i=1}^N \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^N \frac{1}{\sigma_i^2}} = \bar{x}, \quad (4.4)$$

这里 \bar{x} 是各 x_i 以 $\frac{1}{\sigma_i^2}$ 为权重的平均值。 \hat{x} 的期待

值为

$$\langle \bar{x} \rangle = \frac{\sum_{i=1}^N \frac{1}{\sigma_i^2} \langle x_i \rangle}{\sum_{i=1}^N \frac{1}{\sigma_i^2}} = \bar{x}. \quad (4.5)$$

这个结果是预料中的。\$\bar{x}\$ 的方差为

$$\sigma^2(\bar{x}) = \sigma^2(\bar{x}) = \frac{\sum_{i=1}^N \left(\frac{1}{\sigma_i^2}\right)^2 \sigma_i^2}{\left(\sum_{i=1}^N \frac{1}{\sigma_i^2}\right)^2} = \frac{1}{\sum_{i=1}^N \frac{1}{\sigma_i^2}}. \quad (4.6)$$

因此，\$N\$ 组实验结果综合后得到的结果应报道为

$$\bar{x} = \frac{\sum_{i=1}^N \frac{1}{\sigma_i^2} x_i}{\sum_{i=1}^N \frac{1}{\sigma_i^2}} \pm \frac{1}{\sqrt{\sum_{i=1}^N \frac{1}{\sigma_i^2}}}. \quad (4.7)$$

由于各 \$x_i\$ 服从的是正态分布，\$\bar{x}\$ 服从的也是正态分布。故 (4.7) 式给出的是标准的正态分布下的误差报道式。

这个公式非常重要，在实际工作中非常有用。(4.7) 式适用的条件是各实验的结果都服从正态分布，标准误差都是已知的。这个条件实际上一般都会满足。但是要强调的是，在对多个实验的结果进行综合处理时，只通过 (4.7) 式给出报道仍然是不够的，还应当对多个实验的结果作 \$\chi^2\$ 检验。

定义第 \$i\$ 个实验观测值 \$x_i\$ 的 \$\chi^2\$ 量为

$$\chi_i^2 = \frac{(x_i - \bar{x})^2}{\sigma_i^2},$$

其中 \$\bar{x}\$ 为 (4.4) 式所给出的权重平均值，所有实验结果的总 \$\chi^2\$ 量为

$$\chi^2 = \sum_{i=1}^N \chi_i^2 = \sum_{i=1}^N \frac{(x_i - \bar{x})^2}{\sigma_i^2}. \quad (4.8)$$

可以证明这个 \$\chi^2\$ 量服从自由度 \$\nu = N - 1\$ 的 \$\chi^2\$ 分布。由此得出其期待值为 \$N - 1\$，这说明用实验结果计算出的 \$\sum_{i=1}^N \frac{(x_i - \bar{x})^2}{\sigma_i^2}\$ 值应在 \$N - 1\$ 左右摆动。

\$\chi^2\$ 检验就是根据这 \$N\$ 组实验值计算上述 \$\chi^2\$ 量，看其取值是否在 \$N - 1\$ 附近。如果是在 \$N - 1\$ 附近，表明这些实验结果是互相协调的，可以采用 (4.7) 式作为综合后的结果报道。如果计算出的 \$\chi^2\$ 量远大于 \$N - 1\$，则表明这 \$N\$ 组实验结果之间存在不协调。造成这种不协调的原因有三种可能性：(1) 某些实验中有着未被发现的系统误差；(2) 某些实验的误差估计过小，即对实验的测量精度估计过高；(3) 随机的统计涨落。

因此，在遇到有不协调时，一般都应对这些实验重新检查，检查其是否确有被忽视的系统误差，是否对实验的标准误差估计过小。但实际上这一点往往难于做到，而对其是否是由随机的统计涨落造成，则更难于判断。根据这种情况，最常用的一种办法是将所得的标准误差乘一个标度因子

$$S = \left(\frac{\chi^2}{N-1}\right)^{1/2}, \quad (4.9)$$

亦即将结果的报道改为

$$\bar{x} = \frac{\sum_{i=1}^N \frac{1}{\sigma_i^2} x_i}{\sum_{i=1}^N \frac{1}{\sigma_i^2}} \pm \left(\frac{\chi^2}{N-1}\right)^{1/2} \frac{1}{\left(\sum_{i=1}^N \frac{1}{\sigma_i^2}\right)^{1/2}}, \quad (4.10)$$

同时在后面加括号注明 \$S\$ 的取值，以反映这几组实验之间存在着不协调及不协调的程度。例如，高能物理学中测量 \$K^+\$ 介子的平均寿命得到的 \$\tau\$ 值为

\$\tau = (1.2371 \pm 0.0026) \times 10^{-8}\$ 秒 (\$S = 1.9\$)，这个值是由 9 个独立实验结果综合得到的。由于作 \$\chi^2\$ 检验，发现有明显的不协调，因此可按 (4.9) 式和 (4.10) 式给出有标度因子的结果报道。

但是用这种方法报道时，报道的误差也有可能实际上偏大了。因为这种不协调也有可能既不是由系统误差造成，也不是由于对误差估计过小造成，而完全是一种统计涨落。在使用这种报道的结果时应该注意这点。

把标准误差的估计值乘上标度因子 \$S\$，其

物理意义实际上是把原来各实验结果的标准误差都放大 S 倍, 以保证 χ^2 量取值调整到 $N-1$.

$N=2$ 的 χ^2 检验形式特别简单, 这时需要综合的是两个独立实验的结果

$$x = x_1 \pm \sigma_1, \quad x = x_2 \pm \sigma_2.$$

综合后的结果为

$$\bar{x} = \frac{\sigma_2^2 x_1 + \sigma_1^2 x_2}{\sigma_1^2 + \sigma_2^2}, \quad \sigma = \frac{\sigma_1 \sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}. \quad (4.11)$$

χ^2 量为

$$\chi^2 = \frac{(x_1 - x_2)^2}{\sigma_1^2 + \sigma_2^2}. \quad (4.12)$$

χ^2 检验就是看这个 χ^2 是否在 1 附近. 如果它明显大于 1, 则应在报道时加上 S 因子, 这时的

$$S = \frac{|x_1 - x_2|}{\sqrt{\sigma_1^2 + \sigma_2^2}}. \quad (4.13)$$

因此, 当 $(x_1 - x_2)^2 \lesssim \sigma_1^2 + \sigma_2^2$ 时, 两个实验结果综合的报道为

$$x = \frac{\sigma_2^2 x_1 + \sigma_1^2 x_2}{\sigma_1^2 + \sigma_2^2} \pm \frac{\sigma_1 \sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}; \quad (4.14)$$

当 $(x_1 - x_2)^2 > \sigma_1^2 + \sigma_2^2$ 时, 两个实验结果综合的报道为

$$x = \frac{\sigma_2^2 x_1 + \sigma_1^2 x_2}{\sigma_1^2 + \sigma_2^2} \pm |x_1 - x_2| \frac{\sigma_1 \sigma_2}{\sigma_1^2 + \sigma_2^2} \cdot \left(S = \frac{|x_1 - x_2|}{\sqrt{\sigma_1^2 + \sigma_2^2}} \right). \quad (4.15)$$

下面看一个最简单的例子. 化学和物理学中一个常用的常数是阿伏伽德罗 (Avogadro) 常数 N_A , 它的值在 1973 年根据已有的实验确认为

$$N_A = (6.022045 \pm 0.000031) \times 10^{23},$$

现在一般公认的值仍是这个. 但在 1973 年以后, 又有一个实验给出了更精确的结果:

$$N_A = (6.0220978 \pm 0.0000063) \times 10^{23}.$$

按照上面给出的规则, 综合这两个实验结果, 得到

$$\begin{aligned} \bar{N}_A &= 6.0220957 \times 10^{23}, \\ \sigma(\bar{N}_A) &= 0.0000062 \times 10^{23}, \\ \chi^2 &= 2.79 \gg 1. \end{aligned}$$

根据 $\chi^2 \gg 1$, 应该用有标度因子的报道, 结果

为

$$N_A = (6.0220957 \pm 0.0000103) \times 10^{23} \quad (S = 1.7).$$

值得注意的是, 综合后的标准误差本来应该小于每一个实验各自的标准误差, 但由于这两个实验值距离太大而表现出一定的不协调. 为了解决这个困难而引入标度因子后, 标准误差被放大了, 放大到甚致大于其中一个实验值的地步.

从这个例子可以看到, 多个实验结果的综合可以按照上面讲的标准方法进行. 但是在作综合处理时, 应当不要忽视 χ^2 检验. χ^2 检验并不需要进行任何新的实验测量, 只用已有的实验结果就够了, 其作用是检验这几组实验结果之间是否协调, 是否可能存在未被发现的系统误差或对误差估计过小. 因此, 进行 χ^2 检验是对多个实验结果进行综合的不可缺少的组成部分.

二、参数估计的最大似然法

现在从更广的角度讨论参数估计问题的提法和处理方法. 在上一讲中, 我们讨论了直接测量物理量和间接测量物理量的实验结果处理方法, 特别是误差估计问题. 现在要考虑的问题是: 如果需要测量的物理量本身就是某个随机变量概率密度函数中的参数, 这个参数的确定, 原则上需要对与其有关的随机变量进行充分多次测量, 以推算其概率密度函数, 从而得出该参数的估计值. 实际上不可能测量无限多次, 因此需要根据这个随机变量的容量为 N 的样本, 给出该参数的估计值和该估计值的标准误差.

设随机变量 x 服从概率密度函数为 $p(x; \theta_j)$ 的分布, 其中函数 $p(x; \theta_j)$ 的形式是已知的, k 个参数 $\theta_j (j=1, \dots, k)$ 是确定的但却是未知的. 现在对 x 作了 N 次独立测量, 给出了 x 的一个容量为 N 的样本 (x_1, \dots, x_N) , 需要利用这个样本给出参数 $\theta_j (j=1, \dots, k)$ 的估计值及其标准误差.

按照上例的精神,把 N 个 x_i 看作是 N 个独立的随机变量,它们的概率密度函数形式相同,即都是 $p(x; \theta_i)$,并且其中的参数 θ_i 是同一组值.它们的联合概率密度函数为

$$p(x_1, \dots, x_N; \theta_i) = \prod_{i=1}^N p(x_i; \theta_i). \quad (4.16)$$

最大似然法的估计就是选取参数 $\theta_1, \dots, \theta_k$ 的一组值 $\hat{\theta}_1, \dots, \hat{\theta}_k$,使得联合概率密度函数作为 θ_i 的函数在 $\theta_i = \hat{\theta}_i$ 时为极大值.

为简单起见,以后用 $p(x; \theta)$ 代表 $p(x_1, \dots, x_N; \theta_1, \dots, \theta_k)$,只有在涉及特定的 x_i 或 θ_i 时才标 x_i 和 θ_i .这样,(4.16)式可简写为

$$p(x; \theta) = \prod_{i=1}^N p(x_i; \theta). \quad (4.17)$$

注意,等式左边的 p 是联合概率密度函数,它和等式右边的 p 含义不同.从条件概率的意义上来看, $p(x; \theta)$ 可以理解为参数 θ 取值为 θ 时各 x_i 的联合概率密度函数.这样, $p(x; \theta)$ 可以改写为

$$L(x|\theta) = \prod_{i=1}^N p(x_i; \theta), \quad (4.18)$$

$L(x|\theta)$ 称为样本的似然函数. θ 使似然函数取极大值的条件为

$$\frac{\partial}{\partial \theta_i} \ln L(x|\theta) = \sum_{i=1}^N \frac{\partial}{\partial \theta_i} \ln p(x_i; \theta) = 0. \quad (4.19)$$

对应于 $j = 1, \dots, k$,共得到 k 个方程,它们联立解出 $\theta = \hat{\theta}$ 称为 θ 的最大似然估计值.

可以证明:如果用 θ 的函数 $f(\theta)$ 作为参数来找最大似然估计值,只要这函数关系是一一对应的,则所得的最大似然估计值也是一一对应的,也就是

$$\hat{f} = f(\hat{\theta}). \quad (4.20)$$

因此,我们可以根据实际计算的方便,并考虑处理上的简单来选择参数.

问题在于这样得出的参数的估计值的精度如何,即得到的估计值 $\hat{\theta}$ 的标准误差如何.为此,我们先分析 $\hat{\theta}$ 的下列性质:

当样本容量 $N \rightarrow \infty$ 时, $\hat{\theta}$ 的分布趋于一个正态分布,即 $p(\hat{\theta}; \theta) \rightarrow n(\hat{\theta}; \theta, \sigma^2(\theta))$,其方差为

$$\sigma^2(\theta) = \frac{1}{N} \left\langle - \frac{\partial^2}{\partial \theta^2} \ln p(x; \theta) \right\rangle^{-1}. \quad (4.21)$$

注意这里求期待值符号 $\langle \dots \rangle$ 是指对 x 求期待值,因此求期待值后变量 x 实际上不再出现.在实际应用时,常用具体定出的 θ 的估计值 $\hat{\theta}$ 代入(4.21)式中的 θ ,从而给出方差的具体估计值.

下一个问题是在 N 值取多大时上述渐近性质才能足够好地满足,从而可以在较好的精度下来应用.这要取决于概率密度函数 $p(x; \theta)$ 的形式.然而,如果 $p(x; \theta)$ 具有如下的形式:

$$p(x; \theta) = \exp[a(x)\theta + b(x) + c(\theta)], \quad (4.22)$$

则最大似然条件给出

$$\phi(\theta) = - \frac{\partial C(\theta)}{\partial \theta} = \frac{1}{N} \sum_{i=1}^N a(x_i).$$

如果将待估参量取作 $\phi(\theta)$,则有

$$\hat{\phi}(\theta) = \phi(\hat{\theta}) = \frac{1}{N} \sum_{i=1}^N a(x_i), \quad (4.23)$$

$$\sigma^2(\phi(\hat{\theta})) = \frac{1}{N} \left\langle - \frac{\partial^2}{\partial \phi^2} \ln p(x; \theta) \right\rangle^{-1}. \quad (4.24)$$

上式对任意容量的样本都成立.

在实际问题中并不必一定要用这个普遍公式来计算,常常可以根据实际情况采取特殊的,有时往往是更简便的方法.

我们看一个具体的例子:要测定某种不稳定粒子的平均寿命 τ ,已知这种粒子对时间 t 的概率密度函数为

$$p(t; \tau) = \frac{1}{\tau} e^{-t/\tau},$$

如果实际上测量了 N 次,其值为 t_1, \dots, t_N .我们试用几种方法来处理,以资比较.

(1) 用最大似然法处理.首先利用(4.18)式给出样本的似然函数

$$L(t|\tau) = \frac{1}{\tau^N} e^{-\frac{1}{\tau}(t_1 + \dots + t_N)},$$

代入(4.19)式中,由 $\frac{\partial}{\partial \tau} \ln L(t|\tau) = 0$ 得

$$\frac{1}{\tau^2} (t_1 + \cdots + t_N) - N \frac{1}{\tau} = 0.$$

由此得到 τ 的估计值 $\hat{\tau}$ 为

$$\hat{\tau} = \frac{1}{N} (t_1 + \cdots + t_N) \equiv \bar{t},$$

其方差为

$$\begin{aligned} \sigma^2(\tau) &= \frac{1}{N} \left\langle -\frac{\partial^2}{\partial \tau^2} \left(-\frac{t}{\tau} - \ln \tau \right) \right\rangle^{-1} \\ &= \frac{1}{N} \left\langle \frac{2t}{\tau^3} - \frac{1}{\tau^2} \right\rangle^{-1}. \end{aligned}$$

但由于在求期待值时 τ 是常数, 而 $\langle t \rangle = \tau$, 故

$$\sigma^2(\tau) = \frac{1}{N} \left(\frac{1}{\tau^2} \right)^{-1} = \frac{\tau^2}{N}.$$

在对测量结果报道时, 通常将 τ 用估计值 $\hat{\tau}$ 代入. 在 N 的取值很大时, 可近似报道为

$$\tau = \frac{1}{N} \sum_{i=1}^N t_i \left(1 \pm \frac{1}{\sqrt{N}} \right).$$

(2) 在求得了 $\hat{\tau} = \bar{t}$ 之后, 由于 \bar{t} 是一个统计量, 即随机变量 t_1, \cdots, t_N 的函数, 因此求 $\hat{\tau}$ 的方差就归结为求 \bar{t} 的方差. 这样, 直接就得到

$$\begin{aligned} \sigma^2(\hat{\tau}) &= \sigma^2(\bar{t}) = \langle \bar{t}^2 \rangle - \langle \bar{t} \rangle^2 \\ &= \frac{1}{N} (\langle t^2 \rangle - \langle t \rangle^2) \\ &= \frac{\tau^2}{N}. \end{aligned}$$

最后一步是利用熟知的指数分布的结果代入后给出的. 这样得出的报道和上面给出的完全相同.

(3) 在求得了 $\hat{\tau} = \bar{t}$ 之后, 把 τ 的估计值 $\hat{\tau}$ 作为一个 t_1, \cdots, t_N 的函数, 按求统计量概率密度函数的(2.27)式求 $\hat{\tau}$ 的概率密度函数, 得

$$\begin{aligned} p(\hat{\tau}) &= \int \cdots \int \delta \left(\hat{\tau} - \frac{1}{N} \sum_{i=1}^N t_i \right) \\ &\quad \times \frac{1}{\tau^N} e^{-\frac{1}{\tau} \sum_{i=1}^N t_i} dt_1 \cdots dt_N \\ &= N \int \cdots \int \delta \left(N\hat{\tau} - \sum_{i=1}^N t_i \right) \end{aligned}$$

物理

$$\times \frac{1}{\tau^N} e^{-\frac{1}{\tau} \sum_{i=1}^N t_i} dt_1 \cdots dt_N$$

$$= N \frac{1}{\tau^N} \int \cdots \int e^{-N \frac{\hat{\tau}}{\tau}} dt_1 \cdots dt_{N-1}$$

$$N\hat{\tau} - \sum_{i=1}^{N-1} t_i > 0$$

$$= N \frac{\hat{\tau}^{N-1}}{\tau^N} e^{-N \frac{\hat{\tau}}{\tau}} \int \cdots \int dt_1 \cdots dt_{N-1}$$

$$\sum_{i=1}^{N-1} t_i < N$$

$$= \frac{N^N}{(N-1)!} \frac{\hat{\tau}^{N-1}}{\tau^N} e^{-N \frac{\hat{\tau}}{\tau}}$$

这个概率密度自然满足归一化条件, 即

$$\int p(\hat{\tau}) d\hat{\tau} = 1.$$

用这个概率密度函数可导出 $\hat{\tau}$ 的 n 次矩

$$\begin{aligned} \langle \hat{\tau}^n \rangle &= \frac{N^N}{\Gamma(N)} \int \hat{\tau}^n \frac{\hat{\tau}^{N-1}}{\tau^N} e^{-N \frac{\hat{\tau}}{\tau}} d\hat{\tau} \\ &= \frac{N^N}{\Gamma(N)} \tau^n \frac{\Gamma(N+n)}{N^{N+n}} \\ &= \left(\frac{\tau}{N} \right)^n \frac{(N+n-1)!}{(N-1)!}. \end{aligned}$$

由此得到

$$\langle \hat{\tau} \rangle = \tau,$$

$$\sigma^2(\hat{\tau}) = \langle \hat{\tau}^2 \rangle - \langle \hat{\tau} \rangle^2 = \frac{1}{N} \tau^2.$$

将 τ 值用估计值 $\hat{\tau} = \frac{1}{N} \sum_{i=1}^N t_i$ 代入, 当 N 很大时, 可近似得到

$$\tau = \frac{1}{N} \sum_{i=1}^N t_i \left(1 \pm \frac{1}{\sqrt{N}} \right).$$

这种方法是把参数 τ 的估计值 $\hat{\tau}$ 作为样本的函数 (即统计量), 通过样本 t_1, \cdots, t_N 的概率密度函数求出参数估计值 $\hat{\tau}$ 的概率密度函数, 然后求 $\hat{\tau}$ 的各种特征量, 这样得出的结果和前面两种方法得出的完全一致.

在有多多个待估参数的情况下, 也有与上面给出的参数估计值方差的渐近公式相类似的公式, 只不过略为复杂一些.

三、参数估计的置信区间描述

从上面的讨论可以看到，利用最大似然法可以得到参数的估计值，并可以得到此估计值的方差的估计值，通过方差的估计值反映参数估计值的可靠性。但还有两点需要考虑：

(1) 得出参数估计值 $\hat{\theta}$ 的方差 $\sigma^2(\hat{\theta})$ ，其表达式中出现的往往是参数真值 θ 本身。但 θ 是确定的而不是未知的量，因此在具体估算时往往实际上仍用 $\hat{\theta}$ 来代替 θ 。由于 $\hat{\theta}$ 是以 θ 为期望值的随机变量，这个替换又将带来近似性。

(2) 参数 θ 的估计值 $\hat{\theta}$ 服从的分布是由 $p(x; \theta)$ 决定的，它可以是各种形式，常常并不是正态分布。这时如果仅用参数估计值 $\hat{\theta}$ 的标准误差来描述估计值 $\hat{\theta}$ 准确的程度就不够完整了。因为在不同的分布中，在期望值附近一个标准误差内的置信水平是并不相同的。

针对上述情况，在给出参数 θ 的估计值 $\hat{\theta}$ 之后，往往采取置信区间来准确地反映估计值 $\hat{\theta}$ 的准确程度。它的表述形式为

$$P_r(\hat{\theta} - \Delta\theta_1 \leq \theta \leq \hat{\theta} + \Delta\theta_2) = \xi, \quad (4.25)$$

或

$$\theta = \hat{\theta} \begin{matrix} + \Delta\theta_2 \\ - \Delta\theta_1 \end{matrix} \quad (\text{置信水平 } \xi). \quad (4.26)$$

有时在表述形式中不记入参数的估计值 $\hat{\theta}$ ，直接记作

$$P_r(\theta_1 \leq \theta \leq \theta_2 | \theta) = \xi. \quad (4.27)$$

注意这里 θ_1 和 θ_2 是样本 $x_i, i = 1, \dots, N$ 的函数。(4.27) 式的含义是：在 θ 取确定值时，无论它取任何可能的值，它被统计量 θ_1 和 θ_2 所确定的置信区间所包含的概率为 ξ 。

求置信区间的方法很多，需要根据具体情况选择。在上例中已经看到，可以通过样本的概率密度函数求出参数估计值的概率密度函数 $p(\hat{\theta}; \theta)$ 。但是利用 $p(\hat{\theta}; \theta)$ 实际上仍难于作置信区间估计，原因在于参数真值仍作为未知参数进入这个分布，使这个分布实际上还是未定的。因为即使定出了置信区间的上下界，实

际上它们还是 θ 的函数，因而最终还是未知的。

下面介绍两种常用的方法：

1. 利用无参量分布求置信区间

最常用的方法是：设法找一个统计量，它只依赖于样本和待估参数，同时这个统计量又服从一个不包含未知参量的分布，利用这个分布的性质来确定待估参数的置信区间。

首先看上面讨论过的粒子衰变平均寿命 τ 的确定问题。最大似然法给出估计值

$$\hat{\tau} = \bar{t} = \frac{1}{N} \sum_{i=1}^N t_i.$$

上面我们已导出 $\hat{\tau}$ 的概率分布函数为

$$p(\hat{\tau}; \tau) = \frac{N^N}{(N-1)!} \frac{\hat{\tau}^{N-1}}{\tau^N} e^{-\frac{\hat{\tau}}{\tau} N},$$

其中 $\hat{\tau}$ 是上式给出的统计量， τ 是未知的待估计参数。

引入 $\chi^2 = 2 \frac{\hat{\tau}}{\tau} N$ ，代入上式，得到 χ^2 服

从的分布为

$$p(\chi^2) = \frac{1}{2^N (N-1)!} (\chi^2)^{N-1} e^{-\frac{\chi^2}{2}}. \quad (4.28)$$

这正是自由度 $\nu = 2N$ 的 χ^2 分布。由于所选用的统计量 χ^2 只是样本和待估参数的函数，而 χ^2 分布中不再有未知参数，因此符合上述要求，即可以用 χ^2 分布来给出置信区间。

若要求置信水平为 ξ 的置信区间，并将区间上下限取得使在区间两侧外的概率相等，各为 $\frac{1}{2}(1 - \xi)$ ，则置信区间可确定如下：

$$P_r \left(\chi^2_{\frac{1}{2}(1-\xi)}(2N) \leq \chi^2 = 2N \frac{\hat{\tau}}{\tau} \leq \chi^2_{\frac{1}{2}(1+\xi)}(2N) \right) = \xi,$$

或

$$P_r \left(\frac{2N\hat{\tau}}{\chi^2_{\frac{1}{2}(1+\xi)}(2N)} \leq \tau \leq \frac{2N\hat{\tau}}{\chi^2_{\frac{1}{2}(1-\xi)}(2N)} \right) = \xi. \quad (4.29)$$

为了和上一节讲的最大似然法的结果作比较，考虑 $N = 10$ 的情形。要求置信水平 $\xi =$

90%的置信区间,则在(4.29)式中, $N\bar{t} = \sum_{i=1}^{10} t_i$,

查表得

$$\chi_{0.95}^2(20) = 31.4, \chi_{0.05}^2(20) = 10.9.$$

(4.29)式变为

$$P_r(0.64\bar{t} \leq \tau \leq 1.83\bar{t}) = 0.90.$$

这个报道也可改写为

$$\tau = \bar{t} \begin{pmatrix} +0.83 \\ 1.00 \\ -0.36 \end{pmatrix} \quad (\text{置信水平 } 90\%).$$

如果样本容量 $N = 15$, 则结果为

$$\tau = \bar{t} \begin{pmatrix} +0.62 \\ 1.00 \\ -0.31 \end{pmatrix} \quad (\text{置信水平 } 90\%).$$

在用最大似然法求得 $\hat{t} = \bar{t} = \frac{1}{N} \sum_{i=1}^N t_i$

和 $\sigma^2(\hat{t}) = \frac{\tau^2}{N}$ 时, 实际上并不直接给出 \hat{t} 服从的是什么分布. 如果误把它当作正态分布来理解, 并在方差中用 \hat{t} 来代替 τ , 则在 $N = 10$ 和 15 的两种情形下, 置信水平 90% 的报道分别为:

$$\tau = \bar{t}(1.00 \pm 0.52) \quad (\text{置信水平 } 90\%);$$

$$\tau = \bar{t}(1.00 \pm 0.42) \quad (\text{置信水平 } 90\%).$$

这个结果和置信区间描述得到的结果并不一致. 显然, 置信区间描述给出的结果是准确的, 而最大似然法直接求得的只是参数的估计值和方差. 只有在样本容量 N 很大时, 参数的估计值的分布才趋于正态分布. 在上例 $N = 10$ 和 15 时, 当作正态分布给出的置信水平 90% 的报道中, 实际上置信水平约为 87%, 由此可以看出 N 值不够大时带来的影响.

利用无参量分布求置信区间的方法还常用于研究正态分布中的参数估计. 如果随机变量 x 服从正态分布 $n(x; \mu, \sigma^2)$, 对 x 作 N 次测量, 得到期待值 μ 的估计值为 $\hat{\mu} = \bar{x}$. 则可以根据不同情况考虑下列作法:

(1) 因为 \bar{x} 服从 $n(\bar{x}; \mu, \frac{\sigma^2}{N})$, $u = \frac{\bar{x} - \mu}{\sigma/\sqrt{N}}$ 服从 $n(u; 0, 1)$, 如果 σ 已知, 可用 u 来估计 μ 的置信区间;

(2) 如果 σ 未知, 由于 $t = \frac{\bar{x} - \mu}{S_x}$ 服从自

由度为 $N - 1$ 的 t 分布, 可以用 t 来估计 μ 的置信区间;

(3) 由于 $\chi^2 = \frac{1}{\sigma^2} \sum_{i=1}^N (x_i - \bar{x})^2$ 服从自由度为 $N - 1$ 的 χ^2 分布, 可以用 χ^2 来估计 σ^2 的置信区间.

2. 利用似然函数下降求置信区间

似然函数是参数 θ 取值给定时各 x_i 的联合概率密度函数. 在用最大似然法来确定参数的估计值时, 是基于似然函数在给定样本下对参数的估计值取极值的要求. 由此自然想到应能利用似然函数在极值附近随参数估计值的变化来求参数估计值的置信区间. 其作法如下:

如果可以找到适当的参数变换 $\phi = \phi(\theta)$, 使似然函数 $L(x|\theta)$ 的对数可以表为

$$\ln L(x|\theta) = \ln(x|\phi) = -\frac{N}{2} \left(\frac{a(x) - \phi}{\sigma} \right)^2 + b(x), \quad (4.30)$$

则对于任意给定的置信水平 ξ , 可以由

$$\ln L(x|\hat{\theta}) - \ln L(x|\hat{\theta}_{\pm}) = \frac{1}{2} u_{\xi}^2 \quad (4.31)$$

来确定 $\hat{\theta}$ 的置信水平为 ξ 的置信区间 $[\hat{\theta}_-, \hat{\theta}_+]$, 其中 u_{ξ} 为标准正态分布的误差限, 例如:

$$u_{0.683} = 1, \quad u_{0.90} = 1.65, \quad u_{0.95} = 1.96.$$

由此可得出关于参数 θ 估计值的置信区间报道为

$$P_r(\hat{\theta}_- \leq \theta \leq \hat{\theta}_+) = \xi, \quad (4.32)$$

或把 $\hat{\theta}_{\pm}$ 写为 $\hat{\theta}_+ = \hat{\theta} + \delta\hat{\theta}_+$, $\hat{\theta}_- = \hat{\theta} - \delta\hat{\theta}_-$, 而报道写为

$$\theta = \hat{\theta} \begin{pmatrix} +\delta\hat{\theta}_+ \\ -\delta\hat{\theta}_- \end{pmatrix} \quad (\text{置信水平 } \xi). \quad (4.33)$$

采用本方法的关键在于 $\phi(\theta)$ 是否确实存在. 当 N 很大时, 由于似然函数渐近地服从正态分布, 这个条件渐近地被满足. 在这种情况下, $\hat{\theta}$ 的分布趋于正态分布:

$$p(\hat{\theta}; \theta) \rightarrow n(\hat{\theta}; \theta, \sigma^2(\theta)),$$

其中方差 $\sigma^2(\theta)$ 由(4.21)式给出. 在应用这个式子时, 除了要求 N 很大以保证其渐近地服从正态分布外, 主要碰到了两个实际问题: 一个是要把 $-\frac{\partial^2}{\partial \theta^2} \ln p(x; \theta)$ 的期待值求出, 即计

算

$$\left\langle -\frac{\partial^2}{\partial \theta^2} \ln p(x; \theta) \right\rangle = - \int \left[\frac{\partial^2}{\partial \theta^2} \ln p(x; \theta) \right] \times p(x; \theta) dx,$$

这个积分有时并不容易计算；另一个问题是 θ 最终还是以一个未知参量形式出现。针对这两个问题采取的近似作法是：用样本平均代替求期待值的积分，即作代换

$$\langle f(x) \rangle \rightarrow \frac{1}{N} \sum_{i=1}^N f(x_i);$$

然后在最后结果中作代换 $\theta \rightarrow \hat{\theta}$ 。这样， $\sigma^2(\theta)$ 最终完全通过样本 x_1, \dots, x_N 表述出来，即

$$\sigma^2(\theta) \approx \left[-\frac{\partial^2 \ln L(x_1, \dots, x_N | \theta)}{\partial \theta^2} \right]_{\theta = \hat{\theta}(x_1, \dots, x_N)}^{-1}. \quad (4.34)$$

方差给出后，再利用其分布是正态分布，对任意置信水平的置信区间也就都可以给出了。

四、贝斯 (Bayes) 方法和贝斯假设

在对参数估计值及其置信区间的估算问题进行讨论之后，仍可能有不完整的地方。由于参数估计值服从的分布实际上可以是各种不同类型，被估参数的实验估计结果通过估计值和方差（或者置信区间）反映出来总还是不充分的。在参数估计方法中还有另一种方法，即贝斯方法，现作一简要介绍。

1. 参数的验前分布和验后分布

物理量 x 的概率密度函数 $p(x; \theta)$ 实际上的含义为：在参数 θ 取某定值的条件下，随机变量 x 的概率密度函数，即

$$p(x; \theta) = p(x | \theta), \quad (4.35)$$

$$L(x_1, \dots, x_N | \theta) = \prod_{i=1}^N p(x_i | \theta). \quad (4.36)$$

但是在未对 x 进行测量之前，并没有关于 θ 取值的任何具体信息。但这时 θ 作为一个随机变量，实际上应服从一定的概率分布 $p(\theta)$ （尽管可能是未知的），这称为参数 θ 的验前分布。

参数估计要解决的问题实际上是要在实验测得样本 x_1, \dots, x_N 的条件下，取得参数 θ 的

概率密度函数 $p(\theta | x_1, \dots, x_N)$ ，这个函数称为参数 θ 的验后分布。

这个过程可以通过贝斯定理来实现，按照这个定理，有

$$p(\theta | x_1, \dots, x_N) = \frac{p(x_1, \dots, x_N | \theta) p(\theta)}{p(x_1, \dots, x_N)}, \quad (4.37)$$

其中 $p(x_1, \dots, x_N)$ 是 x_1, \dots, x_N 的概率密度函数。

在实际应用时，由于 x_1, \dots, x_N 是实验具体测得的值，因而

$$p(x_1, \dots, x_N | \theta) = \prod_{i=1}^N p(x_i | \theta)$$

是已知的， $p(x_1, \dots, x_N)$ 只是样本的函数，并且其值可由 θ 的验后分布的归一化条件

$$p(x_1, \dots, x_N) = \int p(x_1, \dots, x_N | \theta) p(\theta) d\theta$$

来决定，所以关键的问题是知道参数 θ 的验前分布。

2. 贝斯方法和贝斯假设

如果参数 θ 的验前分布 $p(\theta)$ 已知，则用贝斯定理立刻可得到 θ 的验后分布的全部信息。为了说明这点，仍举一个测量粒子平均寿命的例子。如果这种粒子共测量了 5 个事例，其寿命值分别为 $0.76 \times 10^{-8}s$, $0.43 \times 10^{-8}s$, $1.85 \times 10^{-8}s$, $2.06 \times 10^{-8}s$, $3.10 \times 10^{-8}s$ 。测量的标准误差很小 ($0.02 \times 10^{-8}s$)，因此在作参数估计时可以略去。按前面给出的方法得到平均寿命 τ 的估计值为 $\hat{\tau} = \bar{\tau} = 1.64 \times 10^{-8}s$ ，其标准误差为 $\sigma(\hat{\tau}) = 0.73 \times 10^{-8}s$ 。考虑到它服从的不是正态分布，写出置信水平为 90% 的报道为

$$\tau = \left(\begin{array}{c} 1.64 + 2.52 \\ - 0.74 \end{array} \right) \times 10^{-8}s$$

(置信水平 90%)。

以上结果都是在 τ 是待测参量，并且 τ 取何值并无任何限制的条件得出的。现在我们来看法下情形：如果这种粒子只能是 K 介子或 π 介子，而我们需要根据这两种介子平均寿命的不同来分析实验里观察到的粒子究竟是哪种

介子。

K 介子和 π 介子的平均寿命实验上已经很准确地得到,分别为

$$\tau_K = (1.2371 \pm 0.0026) \times 10^{-8} \text{s},$$

$$\tau_\pi = (2.6030 \pm 0.0023) \times 10^{-8} \text{s}.$$

这样, τ 的验前分布应表现为只能取这两个值中之一。如果没有进一步的物理限制,实际上究竟取哪个值则应该是等概率的。因此有

$$p(\tau) = \frac{1}{2} [\delta(\tau - \tau_K) + \delta(\tau - \tau_\pi)].$$

从上述给定的测量值得到

$$L(t_1, \dots, t_s | \tau) = \frac{1}{\tau^s} e^{-\frac{8.20}{\tau}},$$

其中 t_i 和 τ 都以 10^{-8} 为单位。由贝斯定理得

$$p(\tau | t_1, \dots, t_s) = \frac{\frac{1}{\tau_K^s} e^{-(8.20/\tau_K)} \delta(\tau - \tau_K) + \frac{1}{\tau_\pi^s} e^{-(8.20/\tau_\pi)} \delta(\tau - \tau_\pi)}{\frac{1}{\tau_K^s} e^{-(8.20/\tau_K)} + \frac{1}{\tau_\pi^s} e^{-(8.20/\tau_\pi)}}.$$

以 $\tau_K = 1.24$ 和 $\tau_\pi = 2.60$ 代入,得

$$p(\tau | t_1, \dots, t_s) = 0.56\delta(\tau - 1.24) + 0.44\delta(\tau - 2.60).$$

这表明所观察的粒子有 56% 的概率是 K 介子, 44% 的概率是 π 介子。

从这个例子可以看到, 如果参数的验前分布已知就很好处理了。但问题常常发生在没有关于验前分布的任何知识和信息时, 这时若要用贝斯方法处理就需要引用贝斯假设。

贝斯假设: 如果对参数的验前分布没有任何知识和信息, 则假定它对于一切可取值都是等概率的。

按照贝斯假设, 若 θ 是连续参量, 则 $p(\theta)$ 为均匀分布; 若 θ 是离散型分布参量, 则 $p(\theta)$

对各个离散值都是等概率的。

按照贝斯假设, 参数 θ 的验后分布为

$$p(\theta | x_1, \dots, x_N) = \frac{L(x_1, \dots, x_N | \theta)}{\int L(x_1, \dots, x_N | \theta) d\theta},$$

即似然函数作为 θ 的函数归一化后的结果。

在本文第二节中已经介绍过求粒子平均寿命 τ 的估计值 $\hat{\tau}$ 的方差的几种方法。其中的第三种方法还求出了估计值 $\hat{\tau}$ 的分布, 所得的结果实际上即相当于采用贝斯假设所得的结果。在上面关于粒子是 K 介子还是 π 介子的检验中, 虽然有了参数验前分布的知识, 知道 τ 的可取值只有两个。但验前分布的知识并不完全, 并不知道这两个可取的概率比例。上面采用的验前分布实际上还是用了贝斯假设, 即认为两个可取值的概率是相等的。

贝斯方法和非贝斯方法曾经在数理统计中有过争论。贝斯方法在已知验前分布的情况下处理起来比较简单, 给出的估计可能更精确些。但在验前分布未知的情况下, 非贝斯方法往往更可靠。在没有验前分布知识时用贝斯方法常常引用贝斯假设, 这个假设在一定意义上有相当大的人为性。因为若 ϕ 是 θ 的函数 $\phi = \phi(\theta)$, 若对 θ 引用贝斯假设, 则有

$$p(\theta) d\theta = c d\theta = c \frac{d\theta}{d\phi} d\phi,$$

由此得到 $p(\phi) = c \frac{d\theta}{d\phi}$, 不再是均匀分布了。

由此可见, 在同一问题中, 如果对参数的不同函数引用贝斯假设, 则得到的结果就会完全不同。这个不确定性是引起争论的一个问题, 也是在采用贝斯假设时要谨慎对待的问题



启 事

由中国物理学会表面与界面物理专业委员会举办的 1983 年暑期专题讲习班, 将于 9 月 5 日至 25 日在中国科学院物理研究所举行。