第4章 参数估计

粒子物理实验的目的通常是对于一个或多个感兴趣的物理量进行测量并获得其结果. 假定实验的直接测量对象是随机变量 x, 它的概率密度函数 $f(x;\theta)$ 或 $f(x;\boldsymbol{\theta})$ 中的参数 $\theta(\boldsymbol{\theta})$ 即是我们感兴趣的待测物理量. 对随机变量 x 进行 n 次测量获得 x 的容量 n 的样本 $\boldsymbol{x}=\{x_1,x_2,\cdots,x_n\}$. 数据分析通常是对一个 (或多个) 感兴趣的物理量 θ(或 $\boldsymbol{\theta}$) 应用实验数据样本 $\boldsymbol{x}$ 构建相应的估计量 $\hat{\theta}(\boldsymbol{x})$ 或 $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$ 进行估计. 对于一个特定的实验, 一个物理量 θ(待测参数) 的测量结果通常表示为

$$\theta=\hat{\theta}^{+\sigma_+}_{-\sigma_-}.$$

式中, 估计值 $\hat{\theta}$ 称为参数 θ 的**名义值**或**测量值**; 区间 $\left[\hat{\theta}-\sigma_-,\hat{\theta}+\sigma_+\right]$ 包含参数 θ 真值的概率为 0.6827. 在经典的概率理论中, 参数估计值 $\hat{\theta}$ 的确定属于参数的点估计问题, σ_+,σ_- 的确定属于参数的区间估计问题. 本章讨论参数的点估计, 区间估计问题则留待第 5 章讨论.

4.1 估计量的性质

未知参数 θ 是利用数据样本 $\boldsymbol{x}$ 的某个函数 $\hat{\theta}(\boldsymbol{x})$ 来估计的, $\hat{\theta}$ 称为参数 θ 的**估计量**. 由于随机变量 x 的随机样本 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$ 也是随机变量, 对 $\boldsymbol{x}$ 一组特定的观测值, 估计量 $\hat{\theta}$ 的数值称为**估计值**. 我们对于估计量和估计值将使用同样的记号 $\hat{\theta}$. 一个好的估计量应当具有一致性、无偏性和有效性等品质.

参数估计量的一致性是指: 当观测次数 n 无限增大时, 它的估计值 $\hat{\theta}$ 收敛到参数的真值 θ. 估计量的一致性是样本容量 n 趋于无穷时的极限性质, 而无偏性是估计量在 n 为有限值时的性质, 即要求估计量的数学期望等于未知参数的真值: $E(\hat{\theta})=\theta$. 在估计量为有偏的情形下, 即 $E(\hat{\theta})=\theta+b(\theta)$, 估计量的**偏差** (bias) $b(\theta)$ 应当是已知的或能够用某种方法加以确定.

对于不同的估计量, 方差的大小可以作为它们的有效性的尺度. 当随机变量 x 满足**正规条件**, 即 x 的取值域 (样本空间) 与参数 θ 无关时, 数据样本的联合概率密度 (似然函数)$L(\boldsymbol{x}|\theta)=\prod\limits_{i=1}^{n}f(x_i,\theta)$ 对于 θ 的一阶、二阶导数存在, 则存在参数的一个估计量, 它的方差达到一个最小的下界, 称为**方差下界**, 用符号 MVB (minimum variance bound) 表示. 达到方差下界的估计量称为最小方差估计量, 或**有效估计量**.

方差下界由克拉美-罗 (Cramer-Rao) 不等式给出:

$$\mathrm{MVB} = \frac{1 + \dfrac{\partial b}{\partial \theta}}{I(\theta)}, \tag{4.1.1}$$

式中, $I(\theta)$ 是**费歇尔 (Fisher) 信息量**:

$$I(\theta) = E\left[\left(\frac{\partial \ln L}{\partial \theta}\right)^2\right] = \int \left(\frac{\partial \ln L}{\partial \theta}\right)^2 L\mathrm{d}\boldsymbol{x} = E\left[-\frac{\partial^2 \ln L}{\partial \theta^2}\right] = \int \left(-\frac{\partial^2 \ln L}{\partial \theta^2}\right) L\mathrm{d}\boldsymbol{x}. \tag{4.1.2}$$

一个估计量 $\hat{\theta}$ 的有效性定义为方差下界与该估计量的方差 $V(\hat{\theta})$ 之比: $e(\hat{\theta}) = \mathrm{MVB}/V(\hat{\theta})$. 显然, 我们希望参数的估计量的有效性接近或等于 1. 估计量 $\hat{\theta}$ 的**均方误差** (mean-squre error,MSE) 同时考虑了估计量的方差和偏差导致的不确定性, 它定义为

$$\mathrm{MSE} = E\left[\left(\hat{\theta} - \theta\right)^2\right] = V\left(\hat{\theta}\right) + b^2. \tag{4.1.3}$$

4.2 期望值和方差的估计

期望值和方差是一个随机变量最重要的性质. 假定要根据随机变量 x 的一组测量值 $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$ 来估计 x 的期望值 μ 及其方差 σ^2. x_i $(i = 1, 2, \cdots, n)$ 是 n 个独立的随机变量, 它们有同样的未知期望值 μ 和方差 σ^2. 这相应于在同一个实验中对同一个变量 x 作 n 次独立的测量. 于是 μ, σ^2 的一致、无偏估计为**样本平均** $\bar{x}$ 和**样本方差** S^2:

$$\hat{\mu} = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{4.2.1}$$

$$\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right). \tag{4.2.2}$$

当 x 的期望值 μ 为已知时, 方差 σ^2 的一致、无偏估计量为

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 = \frac{1}{n}\left(\sum_{i=1}^{n} x_i^2 - n\mu^2\right). \tag{4.2.3}$$

它比期望值未知情形下的式 (4.2.2) 给出方差 σ^2 的更好的估计.

$\hat{\mu}$ 的方差为 σ^2/n, 而 $\hat{\sigma}^2$ 的方差为

$$V(\hat{\sigma}^2) = \frac{1}{n}\left(m_4 - \frac{n-3}{n-1}\sigma^4\right), \tag{4.2.4}$$

其中, m_4 是 x 的四阶中心矩.

粒子物理实验测量中, 随机变量 x 往往服从二项分布、泊松分布或正态分布, 容易证明, 这时样本平均是随机变量 x 期望值 μ 的一致、无偏、有效估计量.

当 x 为正态变量时, 式 (4.2.3) 给出期望值 μ 为已知时方差 σ^2 的有效估计, 而样本方差 S^2 给出期望值 μ 为未知时方差 σ^2 的渐近有效估计. 同样, 当 x 为正态变量时, 式 (4.2.4) 变为 $V(\hat{\sigma}^2) = 2\sigma^4/(n-1)$ $(n \geqslant 2)$. 当 n 很大时, $\hat{\sigma}$ 的标准偏差为 $\sigma/\sqrt{2n}$.

如果 x_i $(i = 1, 2, \cdots, n)$ 有不同但已知的误差, 这相应于不同的 n 个实验以不同的误差测量同一个物理量的情形. 假定 x_i 可考虑为正态变量 $N(\mu, \sigma_i^2)$, 则物理量 μ 的无偏估计为**样本加权平均**:

$$\hat{\mu} = \frac{1}{\omega}\sum_{i=1}^{n}\omega_i x_i, \tag{4.2.5}$$

式中, $\omega_i = 1/\sigma_i^2$, $\omega = \sum\limits_i \omega_i$, $\hat{\mu}$ 的标准差为 $1/\sqrt{\omega}$.

4.3 极大似然估计

从数理统计的观点而言, 极大似然法是最重要的参数估计的一般方法, 因为参数的极大似然估计量有许多好的性质.

4.3.1 参数及其方差的极大似然估计

设 x_i $(i = 1, 2, \cdots, n)$ 是随机变量 x 的一组 n 个独立的测量, x 的概率密度函数为 $f(x, \boldsymbol{\theta})$, 其中 $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_k)^{\mathrm{T}}$ 是待定的 k 个未知参数, 则参数 $\boldsymbol{\theta}$ 的**极大似然估计**量 $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$ 是使似然函数 $L(\boldsymbol{x}|\boldsymbol{\theta})$ 达到极大的参数 $\boldsymbol{\theta}$ 的值:

$$L(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta}). \tag{4.3.1}$$

因为似然函数 $L(\boldsymbol{x}|\boldsymbol{\theta})$ 达到极大时其对数 $\ln L$ 也达到极大, 而通常对于 $\ln L$ 求极大的运算比较容易, 所以极大似然估计量可从解**似然方程**求得:

$$\frac{\partial \ln L}{\partial \theta_i} = 0, \quad i = 1, 2, \cdots, k. \tag{4.3.2}$$

极大似然估计量具有**参数变换下的不变性**, 即若参数 $\boldsymbol{\theta}$ 的极大似然估计量为 $\hat{\boldsymbol{\theta}}$, 若选择 $\boldsymbol{\theta}$ 的任意单值函数 $g(\boldsymbol{\theta})$ 作为待估计参数, 令其极大似然估计量为 $\hat{g}(\boldsymbol{\theta})$, 则有 $\hat{g}(\boldsymbol{\theta}) = g(\hat{\boldsymbol{\theta}})$; 这样, 选择 $\boldsymbol{\theta}$ 的任意单值函数 (包括 $\boldsymbol{\theta}$ 自身) 作为待估计参数都能得到 $\boldsymbol{\theta}$ 的相同估计值 $\hat{\boldsymbol{\theta}}$. 同时, 极大似然估计量是渐近无偏的. 当似然函数满足

正规条件时极大似然估计量是一致估计量. 如果参数 $\boldsymbol{\theta}$ 或其函数存在有效估计量, 则它一定是极大似然估计量, 似然方程给出唯一解; 当参数 $\boldsymbol{\theta}$ 或其函数不存在有效估计量时, 则极大似然估计量 $\hat{\boldsymbol{\theta}}$ 具有可能的最小方差. 对于大样本且似然函数满足正规条件的情形, $\hat{\boldsymbol{\theta}}$ 渐近地服从多维正态分布, 其期望值为参数真值, 方差为方差下界 MVB.

参数 $\boldsymbol{\theta}$ 的极大似然估计 $\hat{\boldsymbol{\theta}}$ 只给出参数 $\boldsymbol{\theta}$ 的估计值, 为了了解参数 $\boldsymbol{\theta}$ 的误差, 必须知道 $\hat{\boldsymbol{\theta}}$ 的方差. 对于任意的样本容量 n, 参数 $\hat{\theta}_i$ 和 $\hat{\theta}_j$ 之间的协方差的表达式为

$$V_{ij}(\hat{\boldsymbol{\theta}}) = \int (\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j) L(\boldsymbol{x}|\boldsymbol{\theta}) \mathrm{d}\boldsymbol{x}, \quad i,j = 1,2,\cdots,k. \tag{4.3.3}$$

该积分的计算有时相当困难, 但一般情形下总可以用数值计算求得. 对于极大似然估计 $\hat{\theta}$ 为单个参数的有效估计量的情形, $\hat{\theta}$ 的方差的下述表达式对任意样本容量 n 都是适用的:

$$V(\hat{\theta}) = \frac{\left(1 + \dfrac{\partial b}{\partial \theta}\right)^2}{\left(-\dfrac{\partial^2 \ln L}{\partial \theta^2}\right)_{\theta=\hat{\theta}}}. \tag{4.3.4}$$

特别地, 当 $\hat{\theta}$ 为无偏有效估计量时, 则有

$$V(\hat{\theta}) = \frac{1}{\left(-\dfrac{\partial^2 \ln L}{\partial \theta^2}\right)_{\theta=\hat{\theta}}}. \tag{4.3.5}$$

对样本容量 n 很大和有 k 个未知参数的情形, 若存在 $\boldsymbol{\theta}$ 的 k 个联合、充分统计量, 则极大似然估计 $\hat{\boldsymbol{\theta}}$ 协方差矩阵的逆阵的元素可用下式计算:

$$V_{ij}^{-1}(\hat{\boldsymbol{\theta}}) = \left(-\frac{\partial^2 \ln L}{\partial \theta_i \partial \theta_j}\right)_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \quad i,j = 1,2,\cdots,k. \tag{4.3.6}$$

此外, 对于大样本且似然函数满足正规条件的情形, $\hat{\boldsymbol{\theta}}$ 渐近地服从多维正态分布, 于是有

$$V_{ij}^{-1}(\hat{\boldsymbol{\theta}}) = E\left(-\frac{\partial^2 \ln L(\boldsymbol{x}|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}\right)_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \int \left(-\frac{\partial^2 \ln L(\boldsymbol{x}|\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}\right)_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \cdot L \mathrm{d}\boldsymbol{x}, \quad i,j=1,2,\cdots,k; \tag{4.3.7}$$

或者, 我们可以利用随机变量 x 的概率密度 $f(x,\boldsymbol{\theta})$ 来计算协方差矩阵元:

$$V_{ij}^{-1}(\hat{\boldsymbol{\theta}}) = n \int \frac{1}{n} \left(\frac{\partial f}{\partial \theta_i}\right) \left(\frac{\partial f}{\partial \theta_j}\right) \mathrm{d}x, \quad i,j = 1,2,\cdots,k, \tag{4.3.8}$$

该式的计算不需要随机变量 x 的测量数据, 因而在实验的设计阶段作误差估计特别有用.

如果 x 为正态随机变量, 或样本量 n 充分大, 则似然函数为渐近正态分布, 且 $\ln L$ 为参数 $\boldsymbol{\theta}$ 的二次型函数, 则确定参数 $\boldsymbol{\theta}$ 的 m 倍标准偏差的区域由 $\boldsymbol{\theta}'$ 的边界给出:

$$\ln L(\boldsymbol{\theta}') = \ln L_{\max} - \frac{m^2}{2}, \tag{4.3.9}$$

式中, $L_{\max}$ 是 L 的极大值. 该边界在 θ_i 轴上的投影给出 θ_i 的 m 倍标准偏差**似然区间**. 若 $\ln L$ 不是参数 $\boldsymbol{\theta}$ 的二次型函数, 则上式只能给出近似的 m 倍标准偏差似然区间; 当取 $m=1$ 时, 给出 θ_i 的不对称的正、负误差, 即 $\sigma_+(\theta_i) \neq \sigma_-(\theta_i)$ $(i = 1, 2, \cdots, k)$.

4.3.2 直方图数据的极大似然估计

当随机变量 x 的测量数据个数 n (样本容量) 很大时, 通常表示为**直方图数据**. 当 n 为常数时, 直方图第 i 个子区间中出现 n_i $(i = 1, 2, \cdots, m)$ 个测量值的联合概率密度 (似然函数) 可表示为多项分布

$$L(n_1, n_2, \cdots, n_m | \boldsymbol{\theta}) = n! \prod_{i=1}^{m} \frac{1}{n_i!} p_i^{n_i}. \tag{4.3.10}$$

第 i 个子区间中出现 1 个测量值的概率用 x 的概率密度 $f(x|\boldsymbol{\theta})$ 计算:

$$p_i = p_i(\boldsymbol{\theta}) = \int_{\Delta x_i} f(x|\boldsymbol{\theta}) \mathrm{d}x, \tag{4.3.11}$$

于是似然方程变为

$$\left.\frac{\partial \ln L}{\partial \theta_i}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \frac{\partial}{\partial \theta_i} \left[\sum_{i=1}^{m} n_i \ln p_i(\boldsymbol{\theta})\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0, \quad i = 1, 2, \cdots, m, \tag{4.3.12}$$

求解该组似然方程可得极大似然估计 $\boldsymbol{\theta}$.

4.3.3 广义极大似然估计

若随机变量 x 的测量数据个数 n 不为常数, 而应考虑为期望值为 ν 的随机变量, 则似然函数应为通常的似然函数与观测到 n 个事例的泊松概率之积:

$$L(\nu, \boldsymbol{\theta}) = \frac{\nu^n}{n!} \mathrm{e}^{-\nu} \prod_{i=1}^{n} f(x_i, \boldsymbol{\theta}), \tag{4.3.13}$$

它被称为**广义似然函数**. 参数 $\boldsymbol{\theta}$ 的极大似然估计由下列似然方程的解求得:

$$\frac{\partial \ln L(\nu, \boldsymbol{\theta})}{\partial \theta_j} = 0, \quad j = 1, 2, \cdots, k, \tag{4.3.14}$$

$$\frac{\partial \ln L(\nu, \boldsymbol{\theta})}{\partial \nu} = 0. \tag{4.3.15}$$

当泊松均值 ν 与参数 $\boldsymbol{\theta}$ 无关时, 方程 (4.3.13) 和方程 (4.3.15) 给出 $\hat{\nu}=n$, 方程 (4.3.14) 的解 $\hat{\boldsymbol{\theta}}$ 与方程 (4.3.1) 和方程 (4.3.2) 的解相同. 如果 ν 是参数 $\boldsymbol{\theta}$ 的函数, 则似然方程变为 (略去与 $\boldsymbol{\theta}$ 无关的项)

$$\ln L(\boldsymbol{\theta})=-\nu(\boldsymbol{\theta})+\sum_{i=1}^{n}\ln\left[\nu(\boldsymbol{\theta})\cdot f(x_i,\boldsymbol{\theta})\right]. \tag{4.3.16}$$

由广义似然函数导出的极大似然估计比从一般的似然函数导出的极大似然估计有较小的方差, 因为它同时利用了随机变量 x 的样本测量值 x_i 的信息和样本容量 n 的泊松分布的信息.

对直方图数据, 其广义似然函数为

$$L(n_1,n_2,\cdots,n_m\,|\boldsymbol{\theta})=\prod_{i=1}^{m}\frac{1}{n_i!}\nu_i^{n_i}\mathrm{e}^{-\nu_i}, \tag{4.3.17}$$

其中, ν_i 由 x 的概率密度在 i 子区间内的积分表示:

$$\nu_i=\nu\int_{\Delta x_i}f(x\,|\boldsymbol{\theta})\mathrm{d}x,\quad \nu=\sum_{i=1}^{m}\nu_i. \tag{4.3.18}$$

当泊松期望值 ν 与参数 $\boldsymbol{\theta}$ 无关时, 似然方程变为

$$\left[\frac{\partial\ln L}{\partial\theta_j}\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}=\frac{\partial}{\partial\theta_j}\left[\sum_{i=1}^{m}n_i\ln\nu_i\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}=0,\quad j=1,2,\cdots,m. \tag{4.3.19}$$

它与方程 (4.3.12) 形式相同, 不过 $p_i(\boldsymbol{\theta})$ 用 $\nu_i(\boldsymbol{\theta})$ 代替, 而且 $\hat{\nu}=n$. 如果 ν 是参数 $\boldsymbol{\theta}$ 的函数, 则似然方程变为

$$\left[\frac{\partial\ln L}{\partial\theta_j}\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}=\frac{\partial}{\partial\theta_j}\left[\sum_{i=1}^{m}n_i\ln\nu_i-\nu\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}=0,\quad j=1,2,\cdots,m. \tag{4.3.20}$$

同样, 由广义似然函数导出的极大似然估计比从一般的似然函数导出的极大似然估计有较小的方差, 因为它同时利用了随机变量 x 的样本测量值 x_i 的信息和样本容量 n 的泊松分布的信息.

4.4 最小二乘估计

4.4.1 参数及其方差的最小二乘估计

在 n 个观测点 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$ 通过测量得到一组 n 个观测值 $\boldsymbol{y}=(y_1,y_2,\cdots,y_n)^{\mathrm{T}}$, 相应的观测值真值 $\boldsymbol{\eta}=(\eta_1,\eta_2,\cdots,\eta_n)^{\mathrm{T}}$ 为未知, 它由某个理论

模型预期: $\eta_i = f(x_i, \boldsymbol{\theta})$ $(i = 1, 2, \cdots, n)$, 其中 $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_k)^{\mathrm{T}}$ 是待定的参数. $\boldsymbol{\theta}$ 的**最小二乘估计**可由**最小二乘函数**$Q^2(\boldsymbol{\theta})$ 的极小值求得

$$Q^2(\boldsymbol{\theta}) = (\boldsymbol{y} - \boldsymbol{\eta}(\boldsymbol{\theta}))^{\mathrm{T}} \boldsymbol{V}^{-1} (\boldsymbol{y} - \boldsymbol{\eta}(\boldsymbol{\theta})) = \sum_{i=1}^{n} \sum_{j=1}^{n} (y_i - \eta_i) V_{ij}^{-1} (y_j - \eta_j), \tag{4.4.1}$$

式中, $V_{ij} = \mathrm{cov}(y_i, y_j)$ 是观测值 $\boldsymbol{y}$ 的协方差矩阵 $\boldsymbol{V}$ 的矩阵元. 当 y_i $(i = 1, 2, \cdots, n)$ 为 n 个相互独立的测量时, 最小二乘函数有简单的形式

$$Q^2(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left(\frac{y_i - \eta_i}{\sigma_i} \right)^2, \tag{4.4.2}$$

其中, σ_i 是测量值 y_i 的误差. 一种常见的情形是 y_i 为泊松变量, 则 σ_i^2 等于预期值 η_i, 或可用测量值 y_i 作为近似. 如果 y_i $(i = 1, 2, \cdots, n)$ 为 n 个独立的正态变量, $y_i \sim N(\eta_i, \sigma_i^2)$, 则 $\boldsymbol{y}$ 的似然函数为 $L(\boldsymbol{\theta}) \propto \exp\left[-\frac{1}{2} \sum_{i=1}^{n} \left(\frac{y_i - \eta_i}{\sigma_i} \right)^2 \right]$. 这时$L(\boldsymbol{\theta})$ 对于 $\boldsymbol{\theta}$ 求极大与最小二乘函数 $Q^2(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left(\frac{y_i - \eta_i}{\sigma_i} \right)^2$ 对 $\boldsymbol{\theta}$ 求极小等价, 即参数 $\boldsymbol{\theta}$ 的极大似然估计与最小二乘估计相等.

对于**线性最小二乘估计问题, 即**$f(x_i, \boldsymbol{\theta})$ 是参数 $\boldsymbol{\theta}$ 的线性函数的情形:

$$f(x_i, \boldsymbol{\theta}) = \sum_{j=1}^{k} a_{ij} \theta_j, \quad i = 1, 2, \cdots, n;\ k < n, \tag{4.4.3}$$

其中, a_{ij} 等于 x_i^{j-1} 或 x_i 的 $(j-1)$ 阶勒让德多项式 $(j = 1, 2, \cdots, k)$, 对最小二乘函数 $Q^2(\boldsymbol{\theta})$ 求极小简化为解一组 k 个线性方程. 定义 a_{ij} 为 $n \times k$ 阶矩阵 $\boldsymbol{A}$ 的矩阵元, 对最小二乘函数 $Q^2(\boldsymbol{\theta})$ 求极小得到参数 $\boldsymbol{\theta}$ 的最小二乘估计:

$$\hat{\boldsymbol{\theta}} = (\boldsymbol{A}^{\mathrm{T}} \boldsymbol{V}^{-1} \boldsymbol{A})^{-1} \boldsymbol{A}^{\mathrm{T}} \boldsymbol{V}^{-1} \boldsymbol{y}, \tag{4.4.4}$$

最小二乘估计量 $\hat{\boldsymbol{\theta}}$ 的协方差为

$$\boldsymbol{V}(\hat{\boldsymbol{\theta}}) = (\boldsymbol{A}^{\mathrm{T}} \boldsymbol{V}^{-1} \boldsymbol{A})^{-1}, \tag{4.4.5}$$

或等价地

$$\left(\boldsymbol{V}^{-1}(\hat{\boldsymbol{\theta}}) \right)_{ij} = \left[\frac{1}{2} \frac{\partial^2 Q^2}{\partial \theta_i \partial \theta_j} \right]_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}} = \sum_{l,m=1}^{n} a_{li} a_{mj} (\boldsymbol{V}^{-1})_{lm}, \quad i, j = 1, 2, \cdots, k. \tag{4.4.6}$$

当 $y_i (i = 1, 2, \cdots, n)$ 为 n 个相互独立的测量时, 协方差矩阵 $\boldsymbol{V}$ 为对角矩阵, 非对角矩阵元等于 0, 则式 (4.4.6) 简化为

$$\left(\boldsymbol{V}^{-1}(\hat{\boldsymbol{\theta}}) \right)_{ij} = \sum_{m=1}^{n} a_{mi} a_{mj} / \sigma_m^2, \quad i, j = 1, 2, \cdots, k.$$

线性最小二乘估计量提供了参数的严格解，而且具有理论上的最优性质：唯一性、无偏性和最小方差.

将 $Q^2(\boldsymbol{\theta})$ 函数在极小点 $\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}$ 的邻域作泰勒展开，对于线性模型，在协方差矩阵 $\boldsymbol{V}(\boldsymbol{y})$ 与参数 $\boldsymbol{\theta}$ 无关的条件下，$Q^2(\boldsymbol{\theta})$ 是 $\boldsymbol{\theta}$ 的二次函数，$Q^2(\boldsymbol{\theta})$ 只含两项

$$\begin{aligned}Q^2(\boldsymbol{\theta}) &= Q^2_{\min}+\frac{1}{2}\sum_{i,j}\left(\frac{\partial^2 Q^2}{\partial\theta_i\partial\theta_j}\right)_{\theta=\hat{\theta}}(\theta_i-\hat{\theta}_i)(\theta_j-\hat{\theta}_j)\\ &= Q^2_{\min}+(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})^{\mathrm{T}}\boldsymbol{V}^{-1}(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}).\end{aligned} \tag{4.4.7}$$

当观测值服从期望值为真值的正态分布时，$Q^2(\boldsymbol{\theta})$ 超表面与超平面

$$Q^2(\boldsymbol{\theta})=Q^2(\hat{\boldsymbol{\theta}})+m^2=Q^2_{\min}+m^2 \tag{4.4.8}$$

的截线构成了参数 $\boldsymbol{\theta}=(\theta_1,\theta_2,\cdots,\theta_k)^{\mathrm{T}}$ 的超椭圆联合置信域，该截线在 θ_i 轴上的投影构成 $\hat{\theta}_i$ 的 m 倍标准偏差置信区间. 这时，最小二乘函数的极小值

$$Q^2_{\min}(\boldsymbol{\theta})=\sum_{i=1}^{n}\sum_{j=1}^{n}(y_i-\hat{\eta}_i)V^{-1}_{ij}(y_j-\hat{\eta}_j) \tag{4.4.9}$$

服从自由度 $n-k$ 的 χ^2 分布，这表示最小二乘估计得到的 $Q^2_{\min}$ 是测量值 $\boldsymbol{y}$ 与其拟合值 $\hat{\boldsymbol{\eta}}$ 之间一致性的一种定量表述，即 $Q^2_{\min}$ 代表了拟合的优度.

对于**非线性最小二乘估计问题**，即 $f(x_i,\boldsymbol{\theta})$ 是参数 $\boldsymbol{\theta}$ 的非线性函数的情形，通常对最小二乘函数 $Q^2(\boldsymbol{\theta})$ 求极小需要通过迭代方法来实现，所得到的 $\hat{\boldsymbol{\theta}}$ 只是参数 $\boldsymbol{\theta}$ 的近似值. 非线性**最小二乘估计量是有偏估计量，其方差不可能达到最小方差界**，而且它的 $Q^2_{\min}$ 的分布是未知的. 但是，在样本容量 n 足够大的情形下，**最小二乘估计量是渐近地无偏的**，且其 $Q^2_{\min}$ 近似地服从自由度 $n-k$ 的 χ^2 分布.

4.4.2 直方图数据的最小二乘估计

当随机变量 x 的样本容量 n 很大时，通常表示为直方图数据. 假定直方图第 i 个子区间中出现 $n_i\ (i=1,2,\cdots,m)$ 个测量值，理论模型的相应预期值为

$$f_i(\boldsymbol{\theta})=np_i,\quad p_i(\boldsymbol{\theta})=\int_{\Delta x_i}g(x|\boldsymbol{\theta})\mathrm{d}x, \tag{4.4.10}$$

其中，$g(x|\boldsymbol{\theta})$ 为随机变量 x 的概率密度；$\boldsymbol{\theta}$ 为待定参数. 归一化要求 $\sum\limits_{i=1}^{m}p_i=1$ 意味着

$$\sum_{i=1}^{m}n_i=\sum_{i=1}^{m}f_i(\boldsymbol{\theta})=n. \tag{4.4.11}$$

容易证明, 最小二乘函数 $Q^2(\boldsymbol{\theta})$ 具有如下形式:

$$Q^2(\boldsymbol{\theta})=\sum_{i=1}^{m}\frac{(n_i-np_i)^2}{np_i}=\sum_{i=1}^{m}\frac{(n_i-f_i)^2}{f_i}. \tag{4.4.12}$$

对该最小二乘函数求极小得到参数 $\boldsymbol{\theta}$ 的最小二乘估计, 通常这需要用数值迭代方法求解. 这里 n_i 为均值 np_i 的泊松变量; 当 n 充分大时, n_i 可用期望值和方差均等于 $np_i=f_i$ 的正态变量作为近似. 于是 $(n_i-f_i)/\sqrt{f_i}$ 或 $(n_i-f_i)/\sqrt{n_i}$ 近似地为标准正态变量; 相应地 $Q^2_{\min}(\boldsymbol{\theta})$ 近似地服从 $\chi^2(m-1)$ 分布. 这里自由度为 $m-1$ 是由于约束方程 (4.4.11) 存在.

4.4.3 约束的最小二乘估计

在参数估计问题中, 在观测值 y_i 的真值 $\eta_i\ (i=1,2,\cdots,n)$ 之间常存在一组约束方程. 典型的例子是粒子反应或衰变的运动学分析中能、动量守恒律构成了限制各末态粒子四动量之间的一组约束. 在这样的测量中, 一些物理量以一定的精度加以测定 (如带电径迹的动量和方向), 另一些量则没有加以测量 (如中子、中微子的动量和方向). 现在参数估计的目的是寻找特定的运动学假设所构成的约束条件下的最小二乘估计, 它能给出未测量的物理量的估计以及已测物理量的拟合值, 它是真值的更好的估计.

1) 存在不可测变量

设观测值为 $\boldsymbol{y}=(y_1,y_2,\cdots,y_n)^{\mathrm{T}}$, 测量误差由其协方差矩阵 $\boldsymbol{V}(\boldsymbol{y})$ 表示, 它的真值 $\boldsymbol{\eta}=(\eta_1,\eta_2,\cdots,\eta_n)^{\mathrm{T}}$ 是未知的待估计参数. 此外, 还存在一组 J 个不可直接测量的变量 $\boldsymbol{\xi}=(\xi_1,\xi_2,\cdots,\xi_J)^{\mathrm{T}}$, n 个可测定参数 $\boldsymbol{\eta}$ 和 J 个不可测定参数 $\boldsymbol{\xi}$ 是相关的, 满足一组 K 个约束方程

$$f_k(\boldsymbol{\eta},\boldsymbol{\xi})=0,\quad k=1,2,\cdots,K.$$

要求参数 $\boldsymbol{\eta}$ 和 $\boldsymbol{\xi}$ 的估计值 $\hat{\boldsymbol{\eta}}$ 和 $\hat{\boldsymbol{\xi}}$ 及其误差.

该问题的最小二乘估计求解过程见 3.6.1 节的讨论, 相关的公式见式 (3.6.9)∽式 (3.6.19).

2) 无不可测变量

无不可测变量的运动学约束拟合问题的求解可由 3.6.1 节的推导中去除 J 个不可测定参数 $\boldsymbol{\xi}$ 及其相关项直接得到. 该问题的最小二乘估计求解过程见 3.6.2 节的讨论, 相关的公式见式 (3.6.20)∽ 式 (3.6.28).

3) 约束的最小二乘估计中的自由度

对于线性约束的线性最小二乘估计问题, 并且测量值 $\boldsymbol{y}$ 为多维正态变量, 则 $Q^2_{\min}$ 服从自由度 $K-J$ 的 χ^2 分布; 对于非线性约束的非线性最小二乘估计问题,

或者测量值 $\boldsymbol{y}$ 不是多维正态变量, 当 n 充分大时 $Q^2_{\min}$ 可近似地视为 $\chi^2(K-J)$ 变量.

粒子反应或衰变的运动学分析中, 如果所有末态粒子的径迹参数都作了测量 (没有不测量的参数), 能、动量守恒律构成了限制各末态粒子四动量之间的一组 4 个约束方程, 这时应用约束的最小二乘估计 (称为 4C 运动学拟合) 可得到测量值真值 $\boldsymbol{\eta}$ 的更好的估计值, $Q^2_{\min}$ 可近似地视为 $\chi^2(4)$ 变量. 如果存在 J 个不可直接测量的径迹参数 (如中子、中微子的动量和方向), 并且存在 r 个中间共振态, 它们衰变出的末态粒子的不变质量要求等于母粒子的质量, 这时应用约束的最小二乘估计称为 $(4+r-J)$C 运动学拟合, $Q^2_{\min}$ 可近似地视为 $\chi^2(4+r-J)$ 变量.