粒子物理与核物理实验中的数据分析

陈少敏
清华大学

第九讲：最小二乘法

本讲要点

- 最小二乘法与最大似然法的关系
- 线性情况下的最小二乘估计
- 非线性情况下的最小二乘估计
- 约束情况下的最小二乘法
- 检验最小二乘法的拟合优度
- 应用最小二乘法处理分区数据
- 不等精度关联实验结果的并合问题

最小二乘法与最大似然法

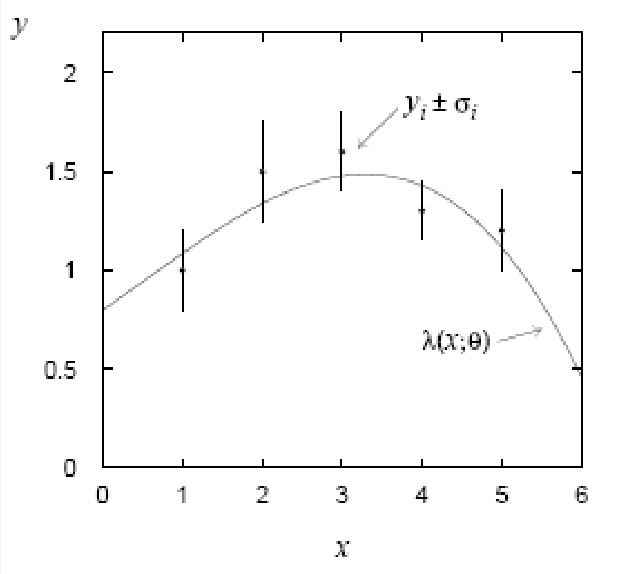
假设有高斯随机变量: y_i, i=1,…,N ,其平均值为

$$E[y_i] = \lambda_i = \lambda(x_i;\vec{\theta})$$

这里，$x_1,\ldots,x_N$ 与 $V[y_i]=\sigma_i^2$ 已知。

为了估计参数$\vec{\theta}$，可以用曲线

拟合所有的测量点(见右图)。



对于独立的高斯变量 y_i，联合概率密度函数为

$$g(\vec{y};\vec{\lambda},\vec{\sigma}^2) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(\frac{-(y_i-\lambda_i)^2}{2\sigma_i^2} \right)$$

求$\log L(\vec{\theta})$的最大值等效于求 χ^2 的最小值。

对应的对数似然函数(去掉与θ 无关的项)为

$$\log L(\vec{\theta}) = -\frac{1}{2}\sum_{i=1}^{N} \frac{[y_i - \lambda(x_i;\vec{\theta})]^2}{\sigma_i^2}$$

$$\chi^2(\vec{\theta}) = \sum_{i=1}^{N} \frac{[y_i - \lambda(x_i;\vec{\theta})]^2}{\sigma_i^2}$$

最小二乘估计量的定义

如果 y_i 是一多维高斯变量，协方差矩阵为V，满足

$$g(\vec{y};\vec{\lambda},V)=\frac{1}{(2\pi)^{N/2}\,|V|^{1/2}}\exp\left[-\frac{1}{2}\left(\vec{y}-\vec{\lambda}\right)^{T}V^{-1}\left(\vec{y}-\vec{\lambda}\right)\right]$$

那么其对数似然函数为

$$\log L(\vec{\theta})=-\frac{1}{2}\sum_{i=1}^{N}[y_i-\lambda(x_i;\vec{\theta})](V^{-1})_{ij}[y_j-\lambda(x_j;\vec{\theta})]$$

也就是说，我们应求下式的最小值

$$\chi^2(\vec{\theta})=\sum_{i,j=1}^{N}[y_i-\lambda(x_i;\vec{\theta})](V^{-1})_{ij}[y_j-\lambda(x_j;\vec{\theta})]$$

它的最小值定义了最小二乘法的估计量 θ，即使 y_i 不是高斯变量，该定义依然适用。(实际上，y_i 通常是高斯的，因为中心极限定理会导出测量误差也是高斯的。)

线性最小二乘法估计

如果$\lambda(x;\vec{\theta})$是$\vec{\theta}$ 的线性函数，最小二乘法有一些简单的特殊性质，

$$\lambda(x;\vec{\theta}) = \sum_{j=1}^{m} a_j(x)\theta_j$$

$\hat{\vec{\theta}}$ 没有偏置，而且得到的方差最小(高斯−马尔可夫定理)。

这里 $a_j(x)$ 是 x 的任意线性独立函数。

用矩阵来表示时，令 $A_{ij}=a_j(x_i)$，有

$$\chi^2(\vec{\theta}) = (\vec{y}-\vec{\lambda})^T V^{-1}(\vec{y}-\vec{\lambda}) = (\vec{y}-A\vec{\theta})^T V^{-1}(\vec{y}-A\vec{\theta})$$

对θ_i 求偏微分，并令结果等于零，有

$$\nabla\chi^2 = -2(A^T V^{-1}\vec{y} - A^T V^{-1}A\vec{\theta}) = 0$$

解方程得到最小二乘法的估计量

$$\hat{\vec{\theta}} = (A^T V^{-1}A)^{-1}A^T V^{-1}\vec{y} \equiv B\vec{y}$$

估计量 $\hat{\theta}_i$ 是测量量 y_i 的线性函数。

非线性最小二乘法估计

如果$\lambda(x;\vec{\theta})$是$\vec{\theta}$ 的非线性函数，最小二乘法不能给出参数的严格解，需要通过迭代法求$\vec{\theta}$的近似解, 使得下式最小

$$\chi^2(\vec{\theta}) = (\vec{y}-\vec{\lambda})^T V^{-1} (\vec{y}-\vec{\lambda})$$

如果采用牛顿法求上式的最小值，第 n+1次迭代公式可采用

$$\hat{\vec{\theta}}^{(n+1)} = \hat{\vec{\theta}}^{(n)} - G^{-1}(\hat{\vec{\theta}}^{(n)}) \bullet g(\hat{\vec{\theta}}^{(n)})$$

$$g_i = \left.\frac{\partial \chi^2}{\partial \theta_i}\right|_{\vec{\theta}=\hat{\vec{\theta}}^{(n)}}, \quad G_{ij} = \left.\frac{\partial^2 \chi^2}{\partial \theta_i \partial \theta_j}\right|_{\vec{\theta}=\hat{\vec{\theta}}^{(n)}}$$

迭代终止判据：

$$\left\| \hat{\vec{\theta}}^{(n+1)} - \hat{\vec{\theta}}^{(n)} \right\| = \left[\sum_{i=1} (\hat{\theta}_i^{(n+1)} - \hat{\theta}_i^{(n)})^2 \right]^{1/2} < \varepsilon$$

ε是一个给定的小数。

$$\hat{\vec{\theta}} = \hat{\vec{\theta}}^{(n+1)}$$

最小二乘估计量的方差

协方差矩阵元$U_{ij} = \mathrm{cov}[\hat{\theta}_i, \hat{\theta}_j]$在线性情况下的误差传递可以写为

$$\hat{\vec{\theta}} = (A^T V^{-1} A)^{-1} A^T V^{-1} \vec{y} \equiv B\vec{y}$$

$$U = BVB^T = (A^T V^{-1} A)^{-1}$$

等效地，可以利用下式来计算

$$(U^{-1})_{ij} = \frac{1}{2}\left[\frac{\partial^2 \chi^2}{\partial\theta_i \partial\theta_j}\right]_{\vec{\theta}=\hat{\vec{\theta}}} = \sum_{k,l}^{N} a_i(x_k)(V^{-1})_{kl}\, a_j(x_l)$$

如果 y_i是高斯变量时，
其与RCF边界一致。

最小二乘估计量的方差(续)

对于$\lambda(x;\vec{\theta})$是参数的线性函数情况下，$\chi^2(\vec{\theta})$是二次型的

$$\chi^2(\vec{\theta}) = \chi^2(\hat{\vec{\theta}}) + \frac{1}{2}\sum_{i,j=1}^{m}\left[\frac{\partial^2 \chi^2}{\partial\theta_i \partial\theta_j}\right]_{\vec{\theta}=\hat{\vec{\theta}}} (\theta_i - \hat{\theta}_i)(\theta_j - \hat{\theta}_j)$$

→ 方差从正切平面到椭圆

$$\chi^2(\vec{\theta}) = \chi^2(\hat{\vec{\theta}}) + 1 = \chi^2_{\min} + 1$$

如果$\lambda(x;\vec{\theta})$不是$\vec{\theta}$的线性函数，上述式子会有偏差，但仍不失为较好的近似。可以把区间$\chi^2(\vec{\theta}) \le \chi^2_{\min} + 1$看作“置信区间”，给出了包含真值$\vec{\theta}$ 的可能性。

注意：上式并不依赖于y_i是否为高斯变量，但无论何种情况，都要计算协方差矩阵$V_{ij} = \mathrm{cov}[y_i, y_j]$。

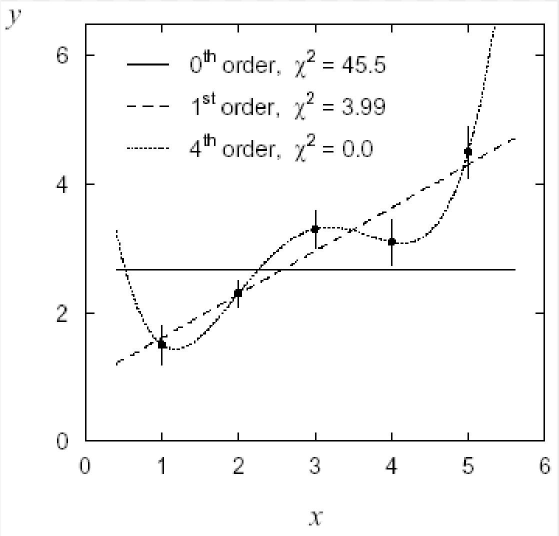
多项式的最小二乘法拟合

用一个多项式来拟合右图

$$\lambda(x;\theta_0,\ldots,\theta_m)=\sum_{j=0}^{m}\theta_j x^j$$

例如:

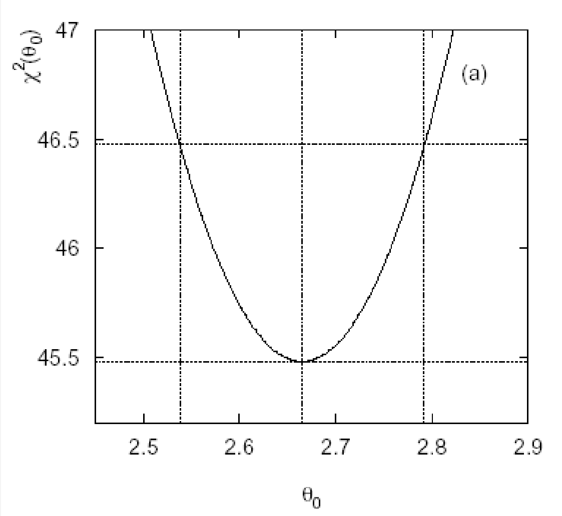
第 0 阶(一个参数)
第 1 阶(两个参数)
第 4 阶(五个参数)



➤对于单参数拟合情况(例如上图的横线):

$$\hat{\theta}_0=2.66\pm0.13$$

$$\chi^2_{\min}=45.5$$



标准误差 $\sigma_{\hat{\theta}_0}$ 是根据 $\chi^2(\hat{\theta}_0\pm\sigma_{\hat{\theta}_0})=\chi^2_{\min}+1$ 来确定的。

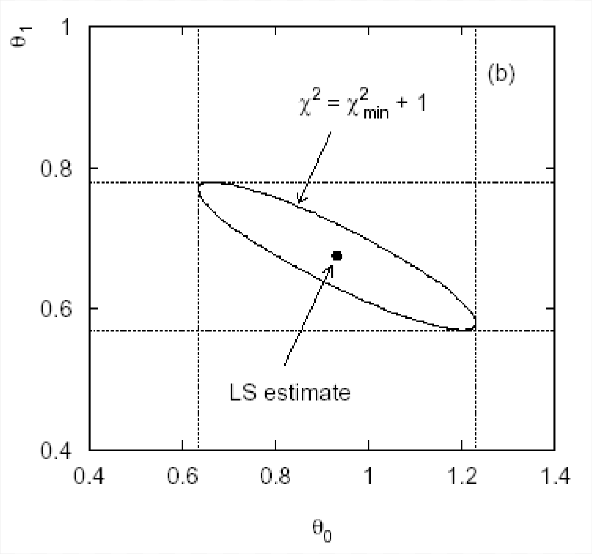
多项式的最小二乘法拟合(续)

➢对于双参数拟合的情形(有非零斜率的直线)

$$\hat{\theta}_0 = 0.93 \pm 0.30; \quad \hat{\theta}_1 = 0.68 \pm 0.10;$$

$$\widehat{\mathrm{cov}}[\hat{\theta}_0, \hat{\theta}_1] = -0.028; \quad r = \hat{\rho} = -0.90;$$

$$\chi^2 = 3.99.$$



切线给出$\sigma_{\hat{\theta}_0}, \sigma_{\hat{\theta}_1}$。

椭圆的倾角给出相关系数(与最大似然法相同)。

虽然表面上可以通过变换$(\hat{\theta}_0, \hat{\theta}_1) \to (\hat{\eta}_0, \hat{\eta}_1)$使得$\mathrm{cov}(\hat{\eta}_0, \hat{\eta}_1) = 0$,从而得到不关联的估计值，但是对新参数的理解也许并不容易。

➢对于五个参量拟合的情形(有非零斜率的直线)

✓曲线通过所有点；

✓$\chi^2_{min} = 0$，参数的数目＝数据点的数目。

χ^2_{min}值的大小反映了数据与假设之间的符合程度。 ➡ 可以用来检验拟合优度。

约束情况下的最小二乘法

在处理实际问题中，常常会遇到测量量本身要受到某些物理定律的约束。例如，能动量守恒，衰变顶点约束等等。无参数的最小二乘问题变为

$$\chi^2(\vec{x}) = (\vec{x}\,' - \vec{x})^T V^{-1} (\vec{x}\,' - \vec{x}) = \text{最小}$$

$$\psi_i(\vec{x}) = 0, \quad i = 1, \ldots, l \quad (\text{共}l\text{个约束条件})$$

$$\vec{x} = (x_1, \ldots, x_m) = \text{真值}$$

$$\vec{x}\,' = (x_1', \ldots, x_m') = \text{观测值}$$

求解可以采用拉格朗日乘子法，对每一个约束引入修正因子α_i，有

$$\chi^2(\vec{x}, \alpha) = (\vec{x} - \vec{x}\,')^T V^{-1} (\vec{x} - \vec{x}\,') + 2\alpha^T \psi(\vec{x}) = \text{最小}$$

可以证明，在经过 **n** 次迭代以后，第 **n**+1 次的值为

$$\alpha^{(n+1)} = \left[F_x(\vec{x}^{(n)}) V F_x^T(\vec{x}^{(n)}) \right]^{-1} \left[\psi(\vec{x}^{(n)}) + F_x(\vec{x}^{(n)})(\vec{x}\,' - \vec{x}^{(n)}) \right]$$

$$\vec{x}^{(n+1)} = \vec{x}\,' - V F_x^T(\vec{x}^{(n)}) \alpha^{(n+1)}, \quad \text{其中 } F_x = \frac{\partial \psi}{\partial x} = l \times m \text{ 矩阵}$$

约束情况下的最小二乘法(续)

当经过 ***n***+1 次迭代以后，满足下式时即可终止

$$\left|\frac{\chi^2(\vec{x}^{(n+1)},\alpha^{(n+1)})-\chi^2(\vec{x}^{(n)},\alpha^{(n)})}{\chi^2(\vec{x}^{(n+1)},\alpha^{(n+1)})}\right|<\varepsilon_1$$

$\left|f(\vec{x}^{(n+1)})\right|<\varepsilon_2$,这里$\varepsilon_1$与$\varepsilon_2$均为小量。

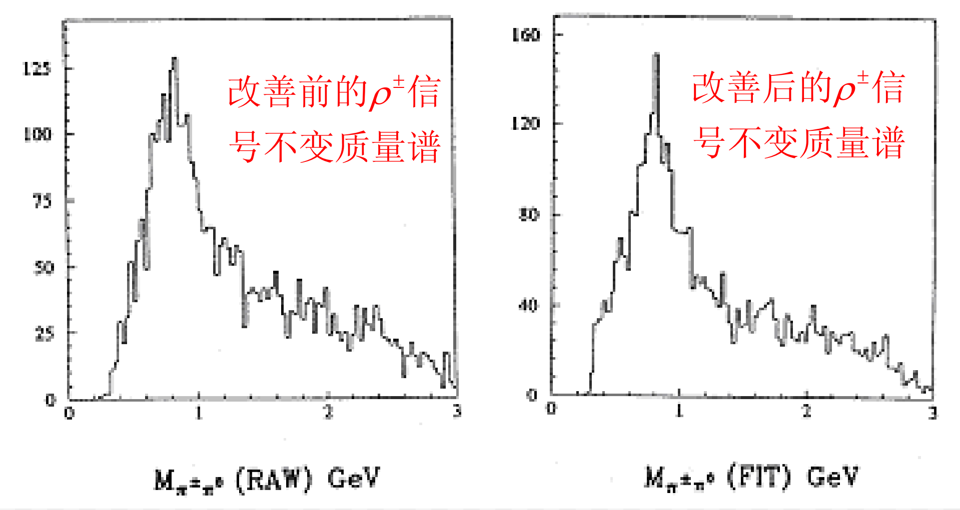
在粒子与核物理实验中，为了提高测量精度而采用的四动量拟合(4-C fit)或顶点拟合(1-C fit)，大都采用该方式来进行。

$$\chi^2_{\min}\sim\chi^2(l)$$

例如，实验观测衰变

$$\rho^{\pm}\rightarrow\pi^{\pm}\pi^{0},\pi^{0}\rightarrow\gamma\gamma$$

对光子的能量测量通常较差，从而影响到 $\rho^{\pm}$ 的重建。可以利用π^0质量约束，进行1-C拟合，改善对光子的测量精度。



检验最小二乘法的拟合优度

如果 $y_i, i=1,\ldots,N$ 是高斯量(V_{ij}已知)，而且假设 $\lambda(x;\vec{\theta})$是θ_i 的线性函数，函数形式正确。

那么$\chi^2_{\min}$ 服从在 N-m自由度下的最小二乘概率密度函数分布。

据此来计算P-值 $P=\int_{\chi^2_{\min}}^{\infty} f(z;n_d)dz$

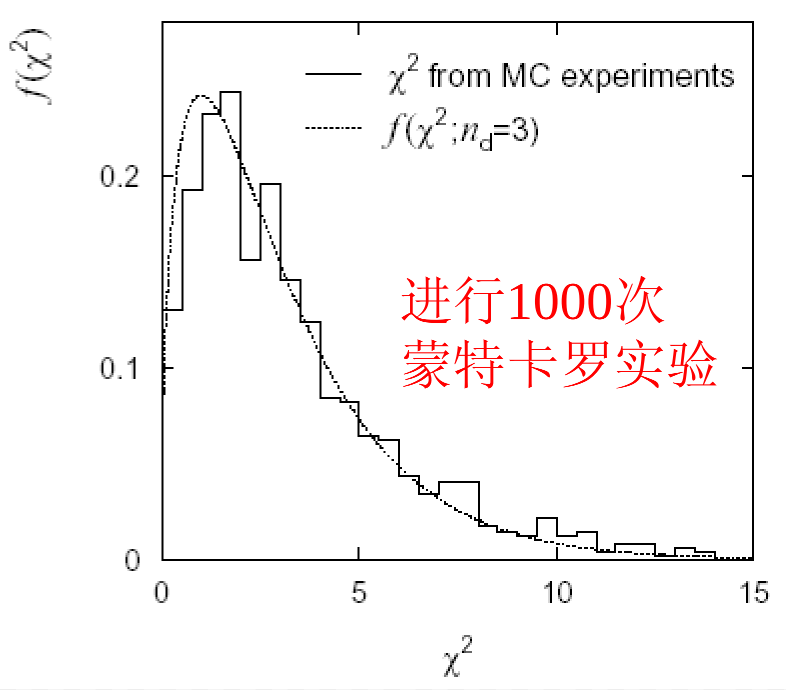
例如考虑在前面双参数拟合的情况

$$\chi^2_{\min}=3.99, N-m=3 \rightarrow P=0.263$$

也就是说，重复实验多次，有26.3%的值将大于 $\chi^2_{\min}$ 。

而对于水平线拟合，有

$$\chi^2_{\min}=45.5, N-m=4 \rightarrow P=3.1\times10^{-9}$$



拟合优度与误差的最小值

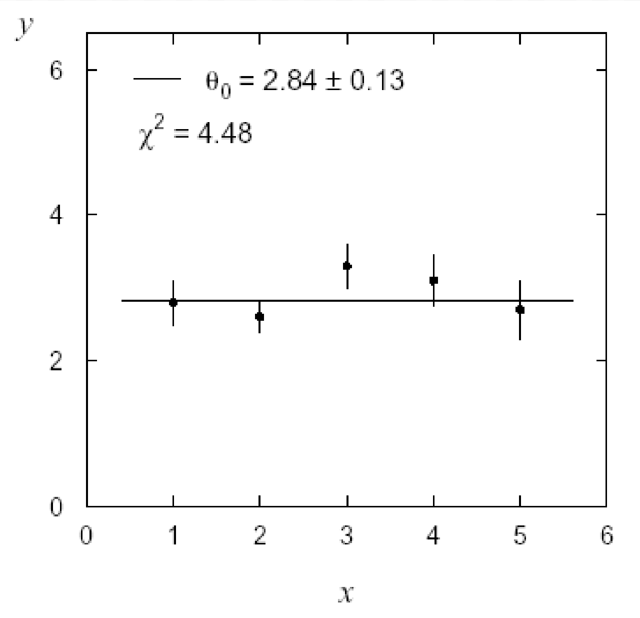
小的统计误差并不意味着是一个好的拟合(反之亦然)

- χ^2曲线在其最小值附近变化给出统计误差；
- χ^2_{min} 的曲率大小给出拟合的优度。

在水平线拟合中，可以人为改变数据点纵向的位置，但保持误差不变，使得

$$\hat{\theta}_0 = 2.66 \pm 0.13 \qquad \chi^2_{\min} = 45.5$$

$$\Rightarrow \quad \hat{\theta}_0 = 2.84 \pm 0.13 \qquad \chi^2_{\min} = 4.48$$



⇒ 方差与改变前一样，但 χ^2_{min} 变“好”了。

⇒ $\chi^2(\theta_0)$曲线只是向下平移，表明与数据符合更好。但曲线形状并没有发生变化，即误差并没有改变。

拟合优度与误差的最小值(续)

- 估计量的方差告诉我们：
  - 如果实验从复多次，估计量 θ 分布有多宽。
  - 但是，它并不告诉我们假设是否正确。
- P-值告诉我们：
  - 如果假设正确，并且实验重复了多次，实验与假设按照统计的χ^2_{min}完全符合或符合得更差的比率是多少。
  - P-值太低，则假设可能有误，即存在系统误差。

最小二乘法处理分区数据

在有 N 个区间，填入 n 次的直方图中，假设概率密度函数为 $f(x;\vec{\theta})$，有

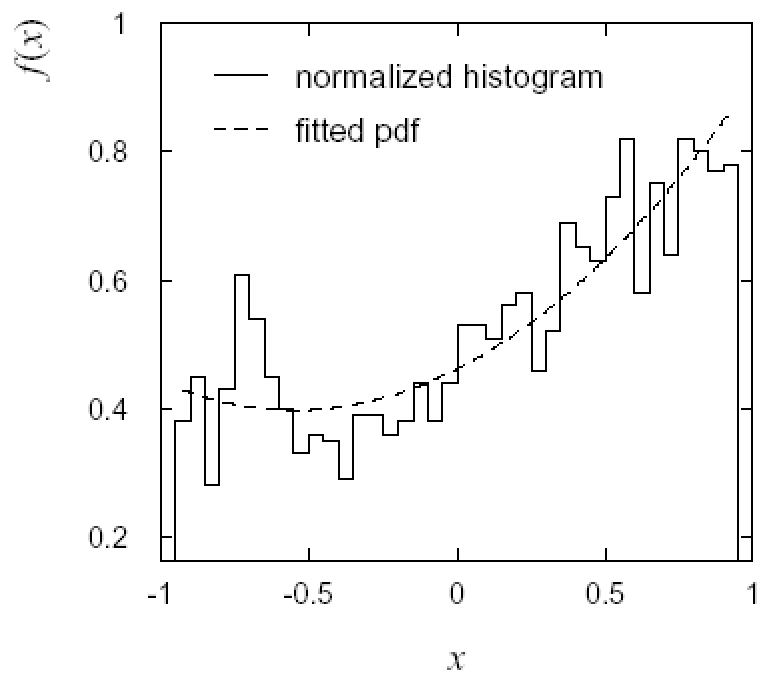
$$y_i = \text{第 } i \text{ 个区间的频数}$$

$$\lambda_i(\vec{\theta}) = n\int_{x_i^{\min}}^{x_i^{\max}} f(x;\vec{\theta})dx = np_i(\vec{\theta})$$

最小二乘法拟合使得下式有最小值

$$\chi^2(\vec{\theta}) = \sum_{i=1}^{N} \frac{(y_i - \lambda_i(\vec{\theta}))^2}{\sigma_i^2}$$

$\sigma_i^2 = V[y_i]$为先验未知量。



把 y_i 看做泊松分布的随机变量，方差为

$$\sigma_i^2 = \lambda_i(\vec{\theta}) \quad \text{(最小二乘法)}$$

$$\sigma_i^2 = y_i \quad \text{(改进后的最小二乘法)}$$

改进的最小二乘法虽方便了计算，但对于有些区间频数太少时$\chi^2_{\min}$不再服从最小二乘的概率密度分布函数(或无定义)。

最小二乘法的归一化问题

尽量避免拟合规一化常数。例如，引入可调参数ν，并与$\vec{\theta}$一起拟合。

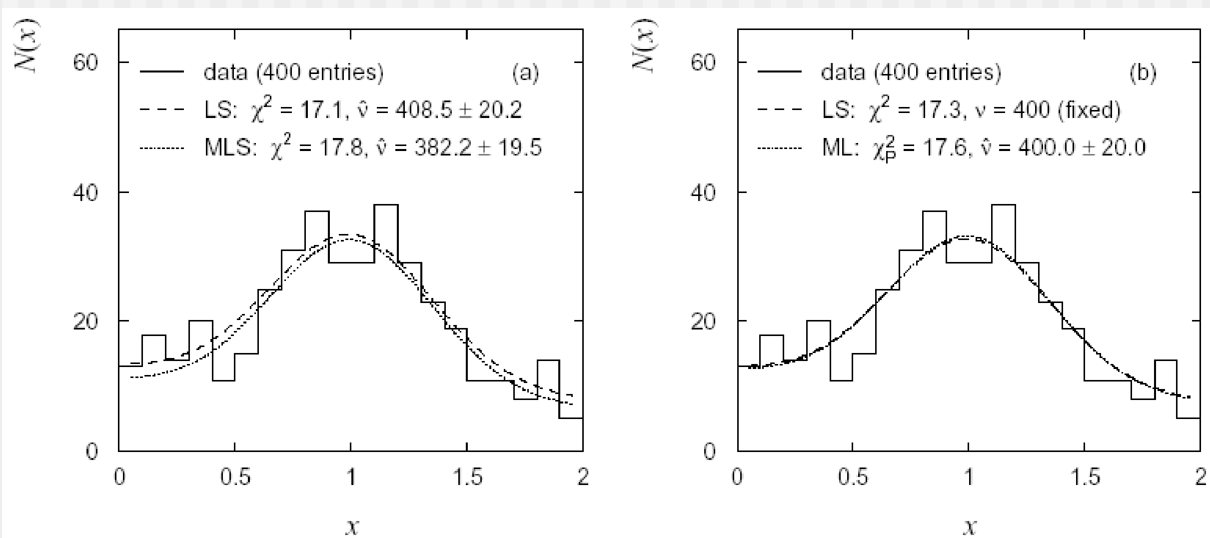
$$\lambda_i(\vec{\theta},\nu) = \nu\int_{x_i^{\min}}^{x_i^{\max}} f(x;\vec{\theta})dx = \nu p_i(\vec{\theta})$$

$\hat{\nu}$ 对 n 而言是一个不好的估计量。

$$\hat{\nu}_{LS} = n + \frac{\chi^2_{\min}}{2}$$

$$\hat{\nu}_{MLS} = n - \chi^2_{\min}$$

例如 n = 400次，N = 20个区间



会出现当 N大 n小时，$\hat{\nu}$ 的相对误差会变大。

解决的方法是从数据中直接得到 n，或者最好是用最大似然法定n。

在PAW上的最小二乘拟合

HISTOGRAM/FIT `id func [ chopt np par step pmin pmax errpar ]`

Only the parameters, which are of more general use, are described in detail. For an up to date description of this command have a look in the online help or in the reference manual.

`ID` A histogram identifier (1-dim or 2-dim)
A bin range may be specified, e.g. `Histo/Fit 10(25:56) ...`

`FUNC` Name of a function to be fitted to the histogram.
This function can be of various forms:

1 The name of a file which contains the user defined function to be minimized. Function name and file name must be the same. For example file `FUNC.FOR` is:

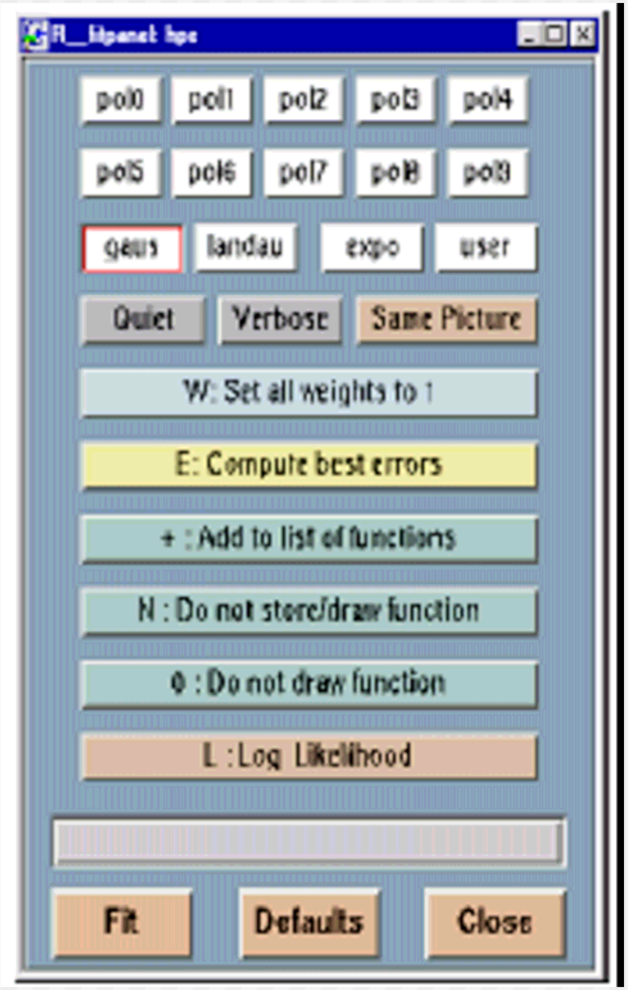
```
FUNCTION FUNC(X)    or FUNC(X,Y) for a 2-Dim histogram
COMMON/PAWPAR/PAR(2)
FUNC=PAR(1)*X  +PAR(2)*EXP(-X)
END
```

2 One of the keywords below (**1-dim histograms only**), which will use the parameterization described at the right for the fit.

```
G    Func=par(1)*exp(-0.5*((x-par(2))/par(3))**2)
E    Func=exp(par(1)+par(2)*x)
Pn   Func=par(1)+par(2)*x+par(3)*x**2...+par(n+1)*x**n, 0<n<20
```

例如: PAW>ve/cr par(3) r 100 0 1;h/fit 100 g ! 3 par (拟合高斯分布)

在ROOT上的最小二乘拟合



在图形显示屏幕中右击鼠标键，可以出现拟合菜单，选择所需的函数对直方图进行拟合。

或者采取如下方式：

root>hist->Fit(“gaus”);
root>hist->Fit(“landau”);
root>hist->Fit(“exp0”);
root>hist->Fit(“pol1”);

…

详见root用户手册

用最小二乘法并合各实验结果

用最小二乘法得到 λ 的 N 个测量的权重平均值

$y_i =$ 第i个测量结果, $i = 1,\ldots,N$

$\sigma_i^2 = V[y_i]$, 假设已知

$\lambda =$ 真值

在各测量量不相关的情况下，并合方法与最大似然法一样。

如果各测量量之间相关, $\mathrm{cov}[y_i, y_j] = V_{ij}$, 通过求下式的最小值

$$\chi^2(\lambda) = \sum_{i,j=1}^{N} (y_i - \lambda)(V^{-1})_{ij}(y_j - \lambda)$$

$$\hat{\lambda} = \sum_{i=1}^{N} w_i y_i, \qquad w_i = \frac{\sum_{j=1}^{N}(V^{-1})_{ij}}{\sum_{k,l=1}^{N}(V^{-1})_{kl}}$$

$$V[\hat{\lambda}] = \sum_{i,j=1}^{N} w_i V_{ij} w_j$$

最小二乘法得到的$\hat{\lambda}$是无偏的, 而且方差最小(高斯–马科夫定理)。

未知关联矩阵的并合问题

$\chi^2(\lambda) > N-1$ 的情形

定义标度因子 → $f = \chi^2/(N-1)$

并合后的误差为 → $\sigma_{\hat{\lambda}} \Rightarrow f\sigma_{\hat{\lambda}}$

粒子数据组PDG建议的处理方法。结果比较保守。

$\chi^2(\lambda) < N-1$ 的情形

对应于多次测量之间存在正关联的情况，得到的误差比真实结果要小。

可以考虑建立等效协方差矩阵(Schmelling M. Phys. Scripta, 1995(51):676)

$$V_{ii} = \sigma_i^2, \quad V_{ij} = f\sigma_i\sigma_j, \quad i \neq j, \quad i,j = 1,\ldots,N$$

解方程

$$\chi^2(f) = \sum_{i,j=1}^{N}(y_i - \hat{\lambda})(V^{-1})_{ij}(y_i - \hat{\lambda}) = N-1 \rightarrow f \rightarrow \text{新的}V[\hat{\lambda}]$$

对两个相关实验求平均的例子

假设有两相关测量量 y_1,y_2，且

$$V=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix} \Rightarrow$$

$$\hat{\lambda}=wy_1+(1-w)y_2,\qquad w=\frac{\sigma_2^2-\rho\sigma_1\sigma_2}{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}$$

$$V[\hat{\lambda}]=\frac{(1-\rho^2)\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2-2\rho\sigma_1\sigma_2}=\sigma$$

如果因第二个测量导致方差倒数的增加为

$$\frac{1}{\sigma^2}-\frac{1}{\sigma_1^2}=\frac{1}{1-\rho}\left(\frac{\rho}{\sigma_1}-\frac{1}{\sigma_2}\right)^2>0 \Rightarrow$$

第二个测量结果对平均值有帮助。

如果$\rho>\sigma_1/\sigma_2\rightarrow w<0$，那么

$\Rightarrow$ 加权平均的结果将不在 y_1 和 y_2 之间。

如果相关是由于使用相同数据的话，上述情况不可能发生，但有却可能来自共同的随机效应；此时的平均值很不可信，例如 ρ，σ_1，σ_2 不正确。

小结

1．与最大似然法的联系

对于高斯分布量 y_i，两者相同

2．线性的最小二乘法估计

估计是通过求矩阵的逆来完成，估计量是测量量 y_i的线性函数

3．非线性的最小二乘法估计

估计是通过迭代来完成，方差可采用线性的情况$\chi^2=\chi^2_{min}+1$来近似

4．用最小二乘法检验拟合优度

用χ^2_{min}做拟合优度统计，满足N-m自由度下的χ^2概率密度函数分布

5．用最小二乘法处理分区数据

把 y_i当作泊松量，得到误差估计λ_i，或y_i(推广最小二乘法)

6．不等精度关联实验结果的并合问题

对关联情况的处理，误差的修正

习题

习题9.1 考虑用最小二乘法拟合直方图。假定直方图分为N个区，以 y_i 表示第 i 个区的频数，预期值为

$$\lambda_i(\vec{\theta}) = n\int_{x_i^{\min}}^{x_i^{\max}} f(x;\vec{\theta})dx$$

这里 $f(x;\vec{\theta})$ 取决于未知参数 $\vec{\theta}$ 。如果用参数来代替总数 n ，同时在对下式进行最小时给出

$$\chi^2(\vec{\theta},\nu) = \sum_{i=1}^{N} \frac{[y_i - \lambda_i(\vec{\theta},\nu)]^2}{\sigma_i^2}$$

a)证明取$\sigma_i^2 = \lambda_i$ ，会导致对总数估计量为

$$\hat{\nu}_{LS} = n + \frac{\chi^2_{\min}}{2}$$

b)并证明若$\sigma_i^2 = y_i$，即改进的最小二乘法，会得到

$$\hat{\nu}_{MLS} = n - \chi^2_{\min}$$

习题(续一)

习题9.2 考虑用最小二乘法拟合直方图。假定直方图分为N个区，以 y_i 表示第 i 个区的频数，预期值为$\lambda_i(\vec{\theta})$。假设总数 n 可以视为常数，从而 y_i 满足多项式分布。

a)协方差矩阵 $V_{ij}=\mathrm{cov}[y_i,y_j]$为何？为什么该矩阵的逆不存在？

b)考虑只拟合头N-1个区的情况。找出协方差矩阵的逆，并证明这等效于在不考虑相关情况下，拟合所有的N个区。

习题9.3 假设一样本容量为 n 的数据样本，每个事例包含了N个测量量 $y_1,\ldots,y_N$，该数据样本将用于拟合一些未知参数。如果测量量之间是相关的，在构造 χ^2量时，需要知道相关矩阵的逆V^{-1}。通常这可以通过首先测定相关系数 $\rho_{ij}=V_{ij}/\sigma_i\sigma_j$来得到，例如，采用蒙特卡罗方法来计算。

a)在RCF边界中，已知有效参数估计量协方差矩阵的逆正比于样本容量 n。证明如果这样的话，相关系数与样本容量无关。

习题(续二)

b)证明协方差矩阵的逆为

$$(V^{-1})_{ij} = \frac{(\rho^{-1})_{ij}}{\sigma_i \sigma_j}$$

(提示：可根据 $\delta_{ij} = \sum_k (V^{-1})_{ik} V_{kj} = \sum_k (V^{-1})_{ik} \rho_{kj} \sigma_k \sigma_j$ ，两边同乘以ρ^{-1}，求和而得到。)

习题9.4 考虑随机变量为x的两个部分重叠样本，分别具有 n 次与m次观测，两者重叠部分为c。假设 x 的方差 $V[x]=\sigma^2$已知，考虑样本平均值为 $y_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$， $y_2 = \frac{1}{m}\sum_{i=1}^{m} x_i$

a)证明协方差为 $\mathrm{cov}[y_1, y_2] = \frac{c\sigma^2}{nm}$

b)求出 y_1 与 y_2 加权平均值和对应的方差或标准误差。