北京大學
PEKING UNIVERSITY

实验物理中的统计方法

第七章：最小二乘法

杨振伟

回顾

- 最大似然法和最大似然估计量
- 四种方法给出最大似然估计的方差
  - ✓ 数值方法
  - ✓ 蒙特卡罗方法
  - ✓ RCF 边界方法
  - ✓ 图解法
- 双参数的最大似然法（等高线）
- 扩展的最大似然法 （样本总量为随机数）
- 最大似然法处理分区数据 （区间大小）
- 用最大似然法合并多组测量结果

本章要点

- 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合并问题

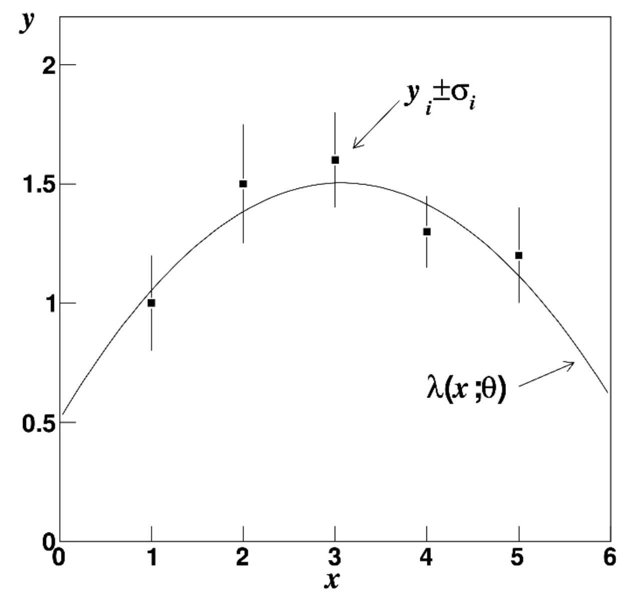
最小二乘法与最大似然法

设高斯变量: $y_i, i = 1, \dots, N$,均值为

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

$x_1, \dots, x_N$ 和 $V[y_i] = \sigma_i^2$ 已知

为估计参数 $\vec{\theta}$，拟合所有测量点



对于独立高斯变量 y_i，联合概率密度为

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

对数似然函数:

$$\ln L(\vec{\theta}) = -\frac{1}{2} \sum_{i=1}^{N} \frac{\left(y_i - \lambda_i\left(x_i; \vec{\theta}\right)\right)^2}{\sigma_i^2} \Rightarrow \chi^2(\vec{\theta}) = \sum_{i=1}^{N} \frac{\left(y_i - \lambda_i\left(x_i; \vec{\theta}\right)\right)^2}{\sigma_i^2}$$

求 $\ln L(\vec{\theta})$ 的最大值等价于求 χ^2 的最小值。

最小二乘估计量

如果 y_i 是多维高斯变量，协方差矩阵为 V，满足

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

对数似然函数为

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

即，我们应求下式的最小值

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

其最小值定义了
最小二乘估计量 $\hat{\theta}$

即使 y_i 不是高斯变量，该定义依然适用。
(实际上，由中心极限定理，y_i 通常是高斯的)

两种情况下的最小二乘参数估计

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

上式对任何含参数函数的具体形式均成立。
实际应用中，可根据理论预期值中所含参数的具体特征而采用不同的参数估计处理方法，简化问题。

线性情况：

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

非线性情况：

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

本章要点

- 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合并问题

线性最小二乘法估计

如果 $\lambda(x;\vec{\theta})$ 是 $\vec{\theta}$ 的线性函数，会有简单的特殊性质：

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

$\hat{\vec{\theta}}$ 没有偏倚，且得到的方差最小
(高斯-马尔科夫定理)

$a_j(x)$：x 的任意线性独立函数

用矩阵来表示，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

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

解方程得到最小二乘估计量

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

估计量 $\hat{\theta}_i$ 是从测量量 y_i 的线性函数。

最小二乘估计量的方差

线性条件下协方差矩阵元$U_{ij} = \mathrm{cov}[\hat{\theta}_i, \hat{\theta}_j]$的误差传递可写为：

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

等效地，可以利用下式来计算

$$\frac{\partial \chi^2}{\partial \vec{\theta}} = -2 \frac{\partial\left(A^T V^{-1} \vec{y} - A^T V^{-1} A \vec{\theta}\right)}{\partial \vec{\theta}} = 2A^T V^{-1} A$$

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

⇒ 如果 y_i 是高斯变量，其方差与RCF边界一致。

最小二乘估计量的方差(续)

对于$\lambda(x;\vec{\theta})$是参数线性函数的情况，$\chi^2(\vec{\theta})$是二次型函数：

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

令 $\chi^2\left(\hat{\vec{\theta}}\right) = \chi^2_{\min}$，上式给出

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

若$\lambda(x;\vec{\theta})$不是$\vec{\theta}$的线性函数，上式有偏差，但仍是好的近似。

可把 $\chi^2(\vec{\theta}) \le \chi^2_{\min} + 1$ 看作“置信区间”给出含真值 $\vec{\theta}$ 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

多项式拟合右图

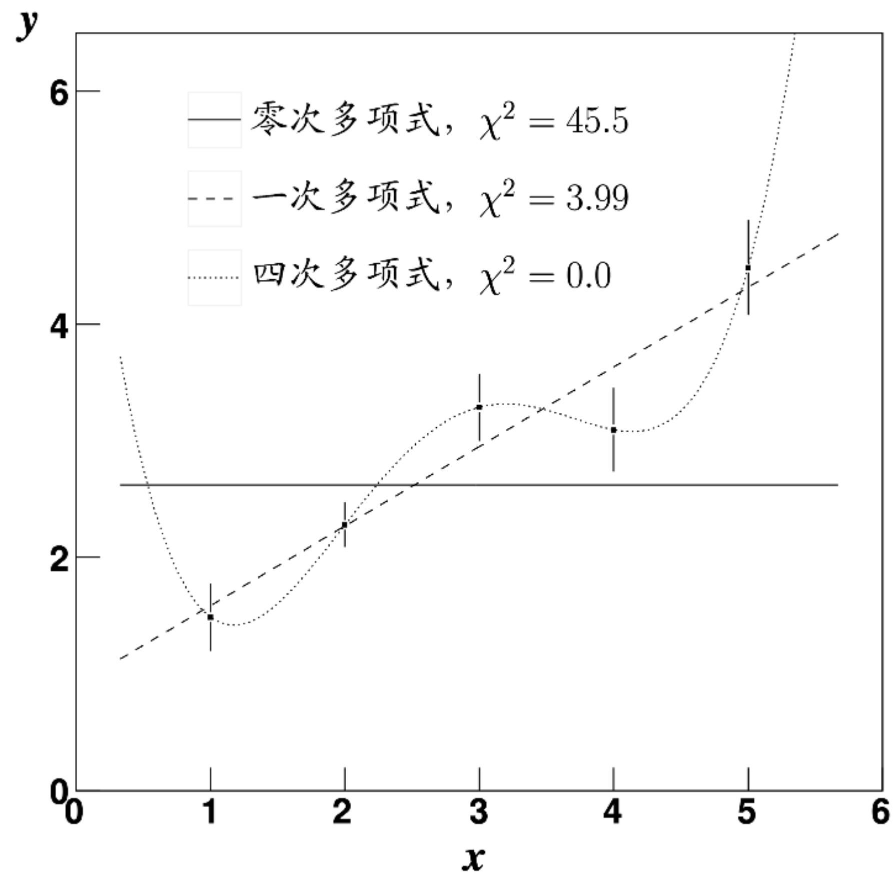
$$\lambda(x;\theta_0,\dots,\theta_m) = \sum_{j=0}^{m} \theta_j x^j$$

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(右图横线):

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



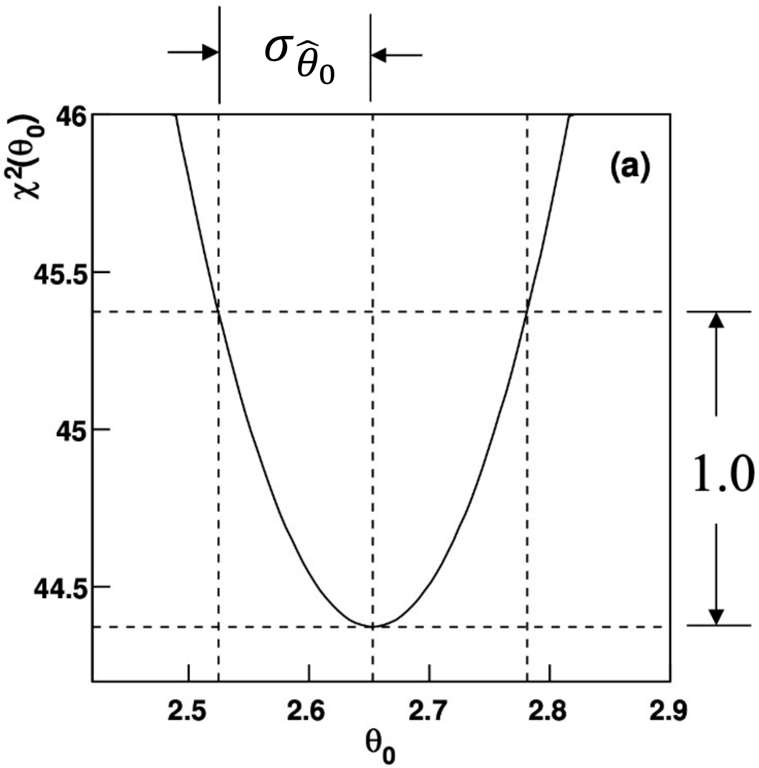
最小二乘估计量的方差

多数情况下与最大似然法中方差估计类似。如果数据服从高斯分布，那么

$$\chi^2(\theta) = -2\ln L(\theta)$$

因此

$$\widehat{\sigma^2}_{\hat{\theta}} \approx 2\left[\frac{\partial^2 \chi^2}{\partial \theta^2}\right]^{-1}_{\theta=\hat{\theta}}$$

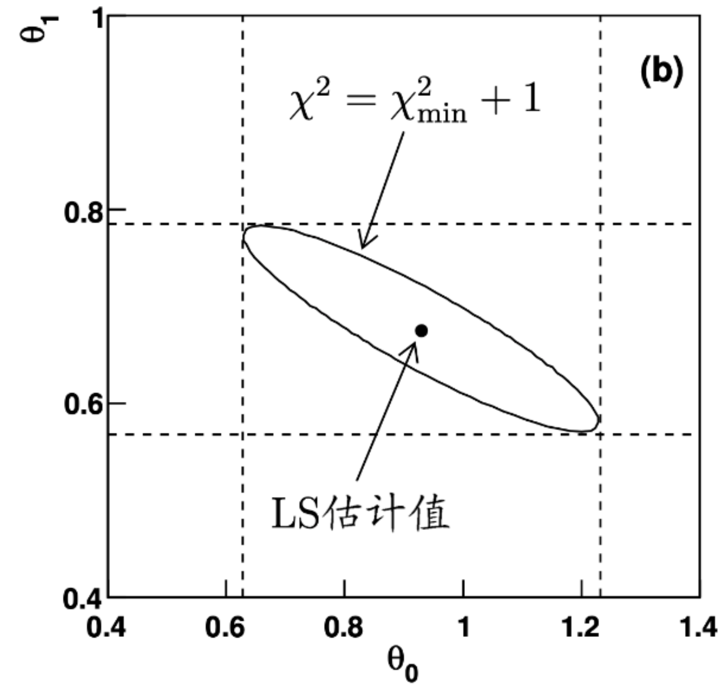


或者，用图解法由$\chi^2(\hat{\theta}_0 \pm \sigma_{\hat{\theta}_0}) = \chi^2_{\min} + 1$确定标准差 $\sigma_{\hat{\theta}_0}$ 。

双参数最小二乘拟合

对于双参数拟合 (以斜率非零的直线为例)

$$\hat{\theta}_0 = 0.93 \pm 0.30$$
$$\hat{\theta}_1 = 0.68 \pm 0.10$$
$$\widehat{\mathrm{cov}}[\hat{\theta}_0, \hat{\theta}_1] = -0.028$$
$$r = \hat{\rho} = -0.90$$
$$\chi^2 = 3.99$$



切线➔$\sigma_{\hat{\theta}_0}, \sigma_{\hat{\theta}_1}$，倾角给出相关系数。

对于五个参数的拟合

- ✓曲线通过所有点;
- ✓$\chi^2_{\min} = 0$，参数个数=数据点个数。

$\chi^2_{\min}$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。

本章要点

- 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合并问题

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，最小二乘法没有参数的解析解，需要通过迭代法求 $\vec{\theta}$ 的近似解，使得下式最小

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

例如，用Minuit（求极小值的程序包）数值求解。

约束情况下的最小二乘法拟合*

有时，测量量本身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。

对一个事例有m个观测量，无参数的最小二乘问题变为

$$\chi^2(\vec{x}) = (\vec{x}' - \vec{x})^T V^{-1} (\vec{x}' - \vec{x}) = 最小$$
$$\psi_i(\vec{x}) = 0,\ i = 1, \dots, l\ \ (共l个约束条件)$$

$$\vec{x}\ = (x_1, \dots, x_m) = 真值$$
$$\vec{x}' = (x_1', \dots, x_m') = 观测值$$

采用拉格朗日乘子法求解，对每一个约束引入因子α_i，

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

实验中，为了提高测量精度而采用四动量守恒约束拟合，顶点或质量约束拟合。

本章要点

- 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合并问题

检验最小二乘法的拟合优度

假设 $y_i, i = 1, \ldots, N$ 是独立的高斯变量（σ_i已知），且 $\lambda(x;\vec{\theta})$ 是θ_i 的线性函数，所采用的函数形式也是正确的，那么

$$\vec{\theta} = \hat{\vec{\theta}} \quad \Rightarrow \quad \chi^2_{\min} = \sum_{i=1}^{N} \frac{\left(y - \lambda\left(x;\hat{\vec{\theta}}\right)\right)^2}{\sigma_i^2}$$

$$\chi^2_{\min} \sim \chi^2(n_\mathrm{d}):\ n_\mathrm{d} = N - m，\ m\text{为参数个数}$$

$\chi^2_{\min}$可以用作拟合优度（goodness-of-fit）统计量，检验假设的函数形式$\lambda(x;\hat{\vec{\theta}})$的好坏。

$\chi^2_{\min}$大 ➔ 拟合不佳（假设的函数形式与数据不符）

$\chi^2_{\min}$小 ➔ 拟合不错（假设的函数形式与数据相符）

$$E[\chi^2(n_\mathrm{d})] = n_\mathrm{d}$$

➔ 如果$\chi^2_{\min} \approx n_\mathrm{d}$则拟合不错

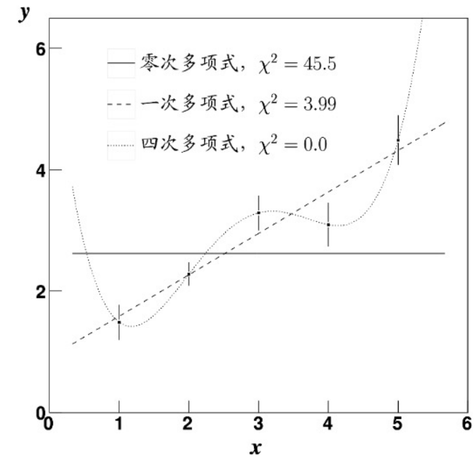
$$p\text{值} = \int_{\chi^2_{\min}}^{\infty} f(z; n_\mathrm{d})\mathrm{d}z$$

检验最小二乘法的拟合优度（续）

例如，前面双参数拟合（直线）

$\chi^2_{\min} = 3.99$，$n_\mathrm{d} = 5 - 2 = 3$
➔ $p = 0.263$

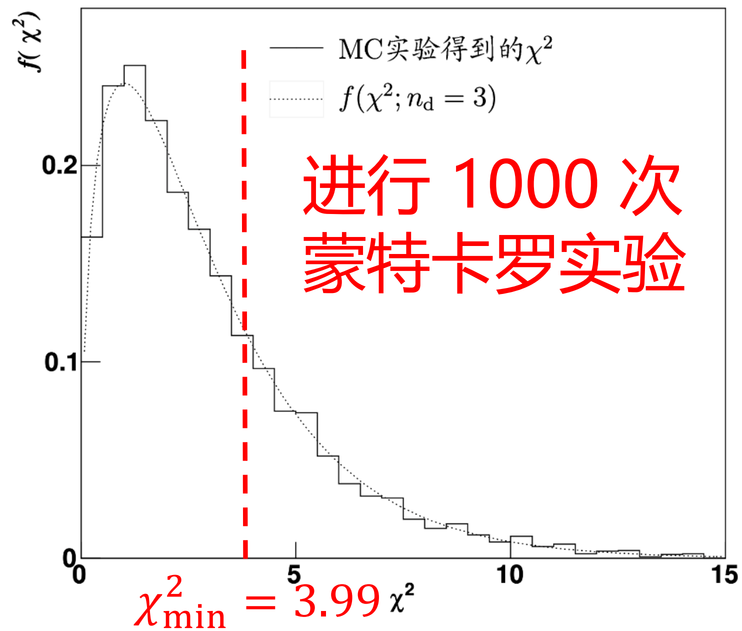
意思是，如果重复实验多次，得到的$\chi^2_{\min}$有26.3% 将不小于3.99。



对于单参数拟合（水平线）：

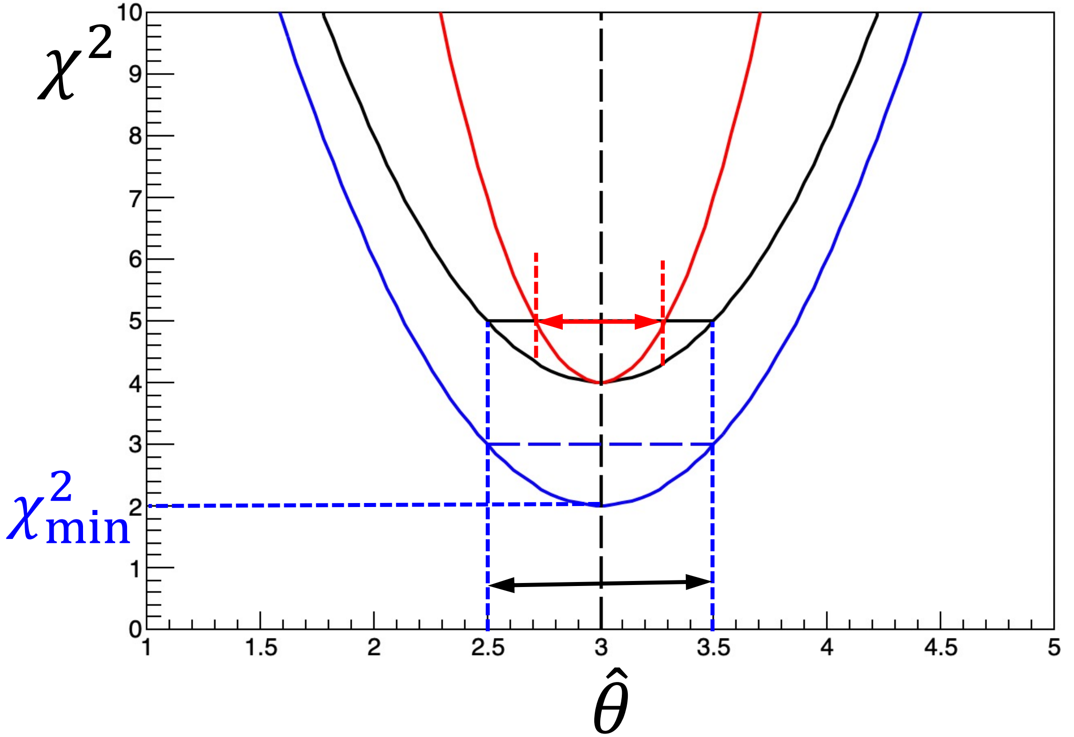
$\chi^2_{\min} = 45.5$，$n_\mathrm{d} = 5 - 1 = 4$
➔ $p = 3.1\times10^{-9}$

p值太小！



拟合优度与统计不确定度的关系

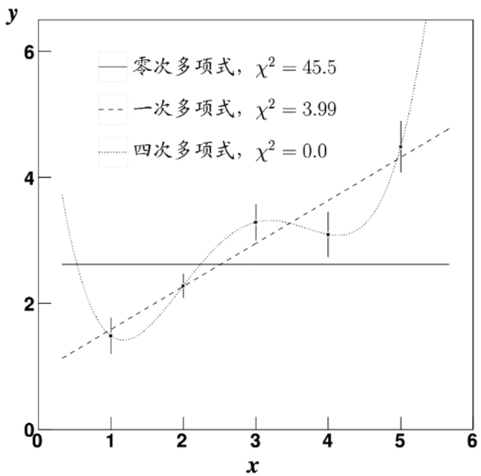
统计不确定度小并不意味着是一个好的拟合(反之亦然)

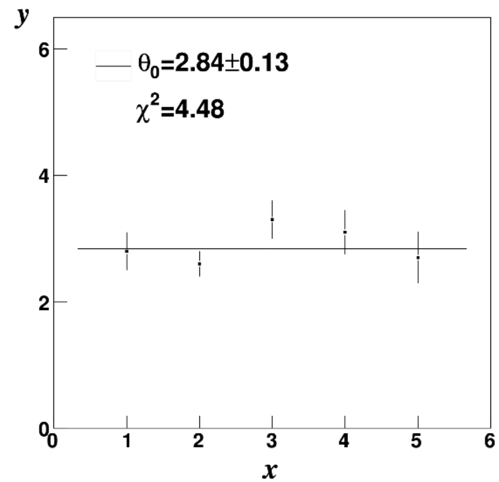


- χ^2曲线在最小值$\chi^2_{\min}$附近变化缓急➔统计不确定度的大小
- $\chi^2_{\min}$的大小➔拟合优度

拟合优度与统计不确定度的关系（续）

水平线拟合中，人为改变数据点纵向位置，但保持误差棒不变





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

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

改变后方差与改变前一样，但 $\chi^2_{\min}$变“好”了。

$\chi^2(\theta_0)$曲线只是向下平移，表明与数据符合更好。但抛物线形状并没有发生变化，即不确定度并没有改变。

拟合优度与不确定度的关系（续）

- 估计量的方差告诉我们:
  - 如果实验从复多次，估计量 $\hat{\theta}$ 的分布有多宽
  - 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，并且重复实验多次，实验和假设符合程度不好于实际结果的统计量$\chi^2_{\min}$的比率是多少
  - p值太低，则假设可能有误，即存在系统不确定度

本章要点

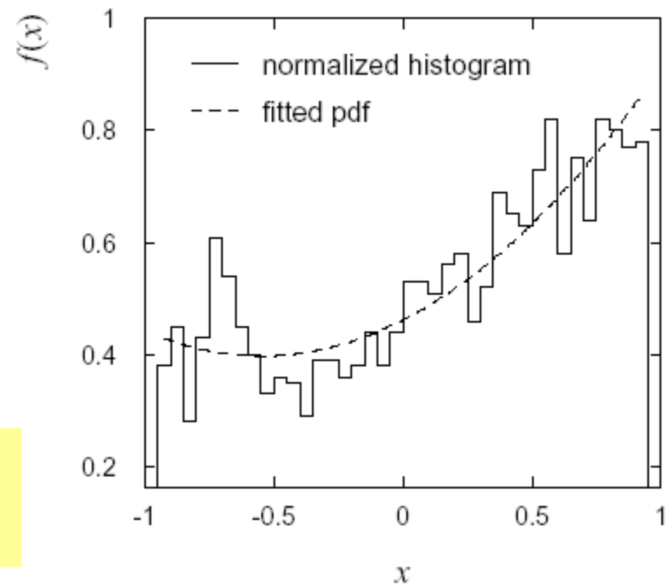
- 最小二乘法与最大似然法的关系
- 线性情况下的最小二乘估计
- 非线性情况下的最小二乘估计*
- 约束情况下的最小二乘法*
- 检验最小二乘法的拟合优度
- 应用最小二乘法处理分区数据
- 不等精度关联实验结果的合并问题

最小二乘法处理分区数据

直方图：N个区间，总频数 n。假设的概率密度函数：$f(x;\vec{\theta})$

y_i = 第 i 个区间的频数

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



最小二乘法拟合使下式有最小值

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

$\sigma_i^2 = V[y_i]$为先验未知量。

把 y_i看作泊松变量，方差为

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

$\sigma_i^2 = y_i$ （改进的最小二乘法MLS）

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

最小二乘法拟合中，尽量避免拟合归一化常数，例如引入可调参数 ν 并与 $\vec{\theta}$ 一起拟合：

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

$\hat{\nu}$ 不是 n 的好估计量，可以证明：

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

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

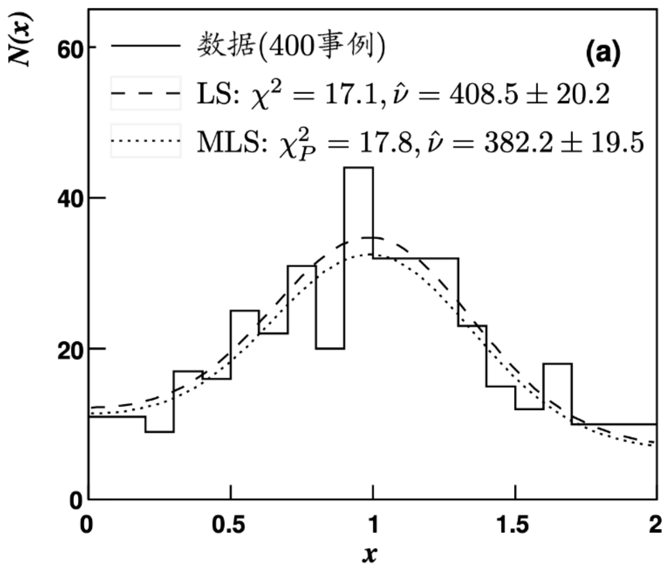
LS和MLS方法得到的估计量$\hat{\nu}$都有偏倚。

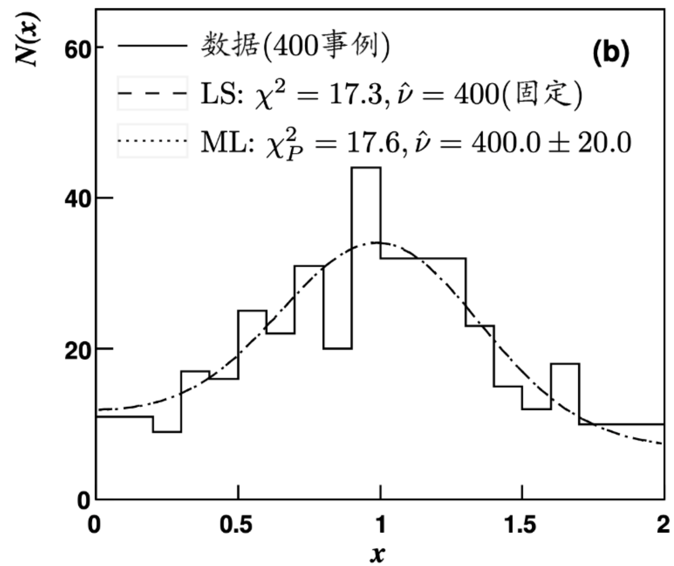
但不管怎样，我们知道偏倚大小。

最小二乘法的归一化问题（续）

例如 $n = 400$次，$N = 20$个区间

$$\hat{\nu}_{\mathrm{LS}} = n + \chi^2_{\min}/2$$
$$\hat{\nu}_{\mathrm{MLS}} = n - \chi^2_{\min}$$





如果 $\chi^2_{\min} \sim n_{\mathrm{d}} = N - m$，那么当 N 较大时，$\hat{\nu}$ 的相对不确定度会比较大。

解决方法：从数据中直接得到 n，或最好采用ML方法确定 n。

本章要点

- 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
- 不等精度相关的实验结果的合并问题

用最小二乘法并合各实验结果

已知 λ 的 N 个测量结果，如何求平均值?

y_i = 第i个测量结果；$\sigma_i^2 = V[y_i]$，假设已知；λ = 真值

如果各测量量之间不相关，求下式最小值：

结果与ML方法一样

$$\chi^2(\lambda) = \sum_{i=1}^{N} \frac{(y_i - \lambda)^2}{\sigma_i^2} \Rightarrow \hat{\lambda} = \sum_{i=1}^{N} \omega_i y_i$$

$$\omega_i = \frac{1/\sigma_i^2}{\sum_{j=1}^{N} 1/\sigma_j^2}, \quad V[\hat{\lambda}] = \sum_{i=1}^{N} \omega_i^2 \sigma_i^2$$

如果各测量量之间相关，$\mathrm{cov}[y_i, y_j] = V_{ij}$，则求下式最小值：

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

$$\omega_i = \frac{\sum_{j=1}^{N} (V^{-1})_{ij}}{\sum_{k,l=1}^{N} (V^{-1})_{kl}}$$

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

LS方法得到的 $\hat{\lambda}$ 是无偏的，且方差最小(高斯-马科夫定理)。

两个相关实验的平均值

假设有两个相关的测量量 y_1和y_2，且

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

$$\hat{\lambda} = w_1 y_1 + (1 - w_1) y_2, \quad w_1 = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

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

因第二个测量导致方差倒数的增加为

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

第二个测量结果对平均值总是有帮助。

如果 $\rho > \sigma_1/\sigma_2 \rightarrow w_1 < 0$，

加权平均的结果将不在 y_1 和 y_2 之间!

如果相关性由使用相同数据引起，不可能发生这种情况；
如果相关性来自共同的随机效应，有可能发生这种情况。
如果ρ, σ_1, σ_2不正确，结果将很不可信，需要自己检查。

例：用不同尺子测量长度

实验上用两把材质不同的尺子测量同一物体的长度。
已知两把尺子在不同温度下有不同的膨胀系数，即

$$y_i = L_i + c_i(T - T_0), \quad i = 1,2$$

根据误差传递公式，可以计算长度估计量的不确定度为

$$\sigma_i^2 = \sigma_{L_i}^2 + c_i^2 \sigma_T^2$$

假设测量是无偏的，即 $E[y_i] = \lambda$

协方差为

$$\begin{aligned} V_{12} &= \mathrm{cov}[y_1, y_2] \\ &= E[y_1, y_2] - \lambda^2 \\ &= c_1 c_2 \sigma_T^2 \end{aligned}$$

协方差矩阵：

$$V = \begin{bmatrix} \sigma_{L_1}^2 + c_1^2 \sigma_T^2 & c_1 c_2 \sigma_T^2 \\ c_1 c_2 \sigma_T^2 & \sigma_{L_2}^2 + c_2^2 \sigma_T^2 \end{bmatrix}$$

例：用不同尺子测量长度（续）

相关系数：

$$\rho = \frac{V_{12}}{\sigma_1\sigma_2} = \frac{c_1c_2\sigma_T^2}{\sqrt{\left(\sigma_{L_1}^2+c_1^2\sigma_T^2\right)\left(\sigma_{L_2}^2+c_2^2\sigma_T^2\right)}}$$

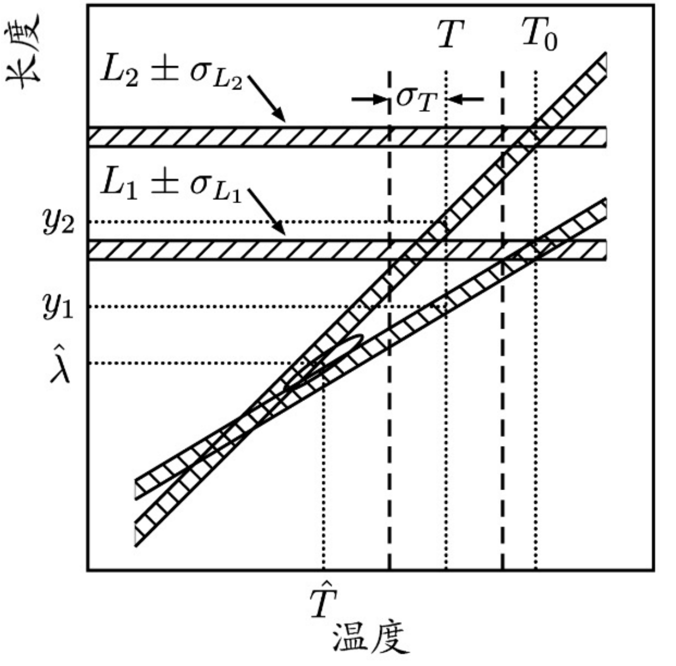
加权平均值：

$$\hat{\lambda} = \frac{\left[\sigma_{L_2}^2+(c_2^2-c_1c_2)\sigma_T^2\right]y_1+\left[\sigma_{L_1}^2+(c_1^2-c_1c_2)\sigma_T^2\right]y_2}{\sigma_{L_1}^2+\sigma_{L_2}^2+(c_1-c_2)^2\sigma_T^2}$$

如果σ_T可以忽略，则 $\hat{\lambda}$ 必在y_1和y_2之间；如果σ_{L_1}和σ_{L_2}可以忽略且$\rho \to 1$，则会出现方差为零的极端情况

$$\hat{\lambda} = \frac{-c_1}{c_1-c_2}y_1 + \frac{c_2}{c_1-c_2}y_2$$

这种情况通常是温度测量极不可靠造成的。



小结

1. 与最大似然法的联系

   对于高斯变量 y_i，二者相同
2. 线性的最小二乘法估计

   通过求矩阵的逆完成估计，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

   通过迭代完成估计，方差可采用线性情况$\chi^2 = \chi^2_{min} + 1$来估计
4. 约束条件下的最小二乘法拟合

   在约束条件下引入拉格朗日乘子改进实验观测量的精度
5. 用最小二乘法检验拟合优度

   用$\chi^2_{\min}$作拟合优度统计，满足$N - m$自由度下的卡方分布
6. 用最小二乘法处理分区数据

   把 y_i 当作泊松变量，不确定用λ_i估计，或用y_i估计 (推广最小二乘法)
7. 不等精度相关的实验结果的并合问题

   对存在相关性的数据的处理，不确定度的修正

似然比与拟合优度

假设用依赖于N个参数$\vec{\mu} = (\mu_1, \dots, \mu_N)$的似然比$L(\vec{\mu})$描述数据。

定义统计量:

$$t_{\vec{\mu}} = -2 \ln \frac{L(\vec{\mu})}{L(\hat{\vec{\mu}})}$$

$\hat{\vec{\mu}}$：$\vec{\mu}$的ML估计量

$t_{\vec{\mu}}$ 的值可以反映假设的 $\vec{\mu}$ 与数据的符合程度:

符合较好意味着 $\hat{\vec{\mu}} \approx \vec{\mu}$ ➔ $t_{\vec{\mu}}$ 的值较小;
$t_{\vec{\mu}}$ 值较大意味着数据与 $\vec{\mu}$ 不太一致。

用 p值定量描述“拟合优度”：

$$p_{\vec{\mu}} = \int_{t_{\vec{\mu},\,\mathrm{obs}}}^{\infty} f(t_{\vec{\mu}}|\vec{\mu})\, \mathrm{d}t_{\vec{\mu}}$$

需要$f(t_{\vec{\mu}}|\vec{\mu})$已知

似然比与拟合优度（续）

假设参数$\vec{\mu} = (\mu_1, \dots, \mu_N)$可以由另一组参数$\vec{\theta} = (\theta_1, \dots, \theta_M)$确定，$M < N$。

例如，在LS拟合中，$\mu_i = \mu\left(x_i; \vec{\theta}\right)$，$x$为控制变量。

定义统计量：

$$q_{\vec{\mu}} = -2\ln\frac{L\left(\vec{\mu}\left(\hat{\vec{\theta}}\right)\right)}{L(\hat{\vec{\mu}})}$$

$\hat{\vec{\theta}}$：拟合M个参数

$\hat{\vec{\mu}}$：拟合N个参数

用 $q_{\vec{\mu}}$ 检验假设的函数形式 $\mu\left(x; \vec{\theta}\right)$。

要得到p值，需要知道$f\left(t_{\vec{\mu}} | \vec{\mu}\right)$。

Wilks 定理（1938）

Wilks 定理：
如果假设的参数$\vec{\mu} = (\mu_1, \ldots, \mu_N)$为真，那么在大数据样本极限下（同时满足其他一些条件），$t_{\vec{\mu}}$和$q_{\vec{\mu}}$服从卡方分布。

当分子中的 $\vec{\mu} = (\mu_1, \ldots, \mu_N)$ 固定时：

$$t_{\vec{\mu}} = -2\ln\frac{L(\vec{\mu})}{L(\hat{\vec{\mu}})}$$

$$f\left(t_{\vec{\mu}}\middle|\vec{\mu}\right) \sim \chi^2(N)$$

如果分子中有M个可调参数：

$$q_{\vec{\mu}} = -2\ln\frac{L\left(\vec{\mu}\left(\hat{\hat{\vec{\theta}}}\right)\right)}{L(\hat{\vec{\mu}})}$$

$$f\left(q_{\vec{\mu}}\middle|\vec{\mu}\right) \sim \chi^2(N-M)$$

S.S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, Ann. Math. Statist. **9** (1938) 60-2.

高斯数据的拟合优度

假设数据是N个独立的高斯分布的值：

$$y_i \sim N\left(\mu_i, \sigma_i^2\right), \quad i = 1, \dots, N$$

μ_i：待估计参数
σ_i^2：已知参数

似然函数：

$$L(\vec{\mu}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp - \frac{(y_i-\mu_i)^2}{2\sigma_i^2}$$

ML估计量：

$$\hat{\mu}_i = y_i$$
$$i = 1, \dots, N$$

对数似然函数：

$$\ln L(\vec{\mu}) = -\frac{1}{2}\sum_{i=1}^{N} \frac{(y_i-\mu_i)^2}{\sigma_i^2} + C$$

拟合优度统计量：

$$t_{\vec{\mu}} = -2\ln\frac{L(\vec{\mu})}{L(\hat{\vec{\mu}})} = \sum_{i=1}^{N} \frac{(y_i-\mu_i)^2}{\sigma_i^2}$$

$$f(t_{\vec{\mu}}|\vec{\mu}) \sim \chi^2(N)$$

$$t_{\vec{\mu}} = -2\ln\frac{L(\vec{\mu}(\hat{\vec{\theta}}))}{L(\hat{\vec{\mu}})} = \sum_{i=1}^{N} \frac{\left(y_i-\mu_i(\hat{\theta})\right)^2}{\sigma_i^2}$$

$$f(q_{\vec{\mu}}|\vec{\mu}) \sim \chi^2(N-M)$$

所以，Wilks 定理形式上给出了LS拟合得到的最小卡方的著名性质。

泊松数据的拟合优度

假设数据$\vec{n} = (n_1, \dots, n_N)$是独立的泊松分布值：

$n_i \sim \pi(\nu_i), \quad i = 1, \dots, N$

$\vec{\nu} = (\nu_1, \dots, \nu_N)$：待估计参数

似然函数：$L(\vec{\nu}) = \prod_{i=1}^{N} \frac{\nu_i^{n_i}}{n_i!} e^{-\nu_i}$

对数似然函数：$\ln L(\vec{\nu}) = \sum_{i=1}^{N} (n_i \ln \nu_i - \nu_i) + C$

ML估计量：

$\hat{\nu}_i = n_i$
$i = 1, \dots, N$

拟合优度统计量：

$$t_{\vec{\nu}} = -2 \ln \frac{L(\vec{\nu})}{L(\hat{\vec{\nu}})} = -2 \sum_{i=1}^{N} \left[n_i \ln \frac{\nu_i}{n_i} - \nu_i + n_i \right]$$

$$q_{\vec{\nu}} = -2 \ln \frac{L\left(\vec{\nu}(\hat{\vec{\theta}})\right)}{L(\hat{\vec{\nu}})} = -2 \sum_{i=1}^{N} \left[n_i \ln \frac{\nu_i(\hat{\vec{\theta}})}{n_i} - \nu_i(\hat{\vec{\theta}}) + n_i \right]$$

Wilks 定理 $f(t_{\vec{\nu}} | \vec{\nu}) \sim \chi^2(N)$ $\quad f(q_{\vec{\nu}} | \vec{\nu}) \sim \chi^2(N - M)$

利用 t和q 量化拟合优度（p值），可以利用Wilks定理抽样其分布。
在大样本极限下严格成立，在小样本的情况下也是很好的近似。

多项分布数据的拟合优度

假设数据$\vec{n} = (n_1, \ldots, n_N)$是多项分布的值：

$$P(\vec{n}|\vec{p}, n_{\text{tot}}) = \frac{n_{\text{tot}}!}{n_1!\, n_2! \cdots n_N!} p_1^{n_1} p_2^{n_2} \cdots p_N^{n_N}$$

$n_{\text{tot}} = \sum_{(i=1)N} n_i$

对数似然函数：$\ln L(\vec{\nu}) = \sum_{i=1}^{N} n_i \ln \frac{\nu_i}{n_{\text{tot}}} + C$

$\nu_i = p_i n_{\text{tot}}$

ML估计量：$\boldsymbol{\hat{\nu}_i = n_i}$ 只有$N-1$个独立量

拟合优度统计量：

$$t_{\vec{\nu}} = -2 \ln \frac{L(\vec{\nu})}{L(\hat{\vec{\nu}})} = -2 \sum_{i=1}^{N} \left[n_i \ln \frac{\nu_i}{n_i} \right]$$

$$q_{\vec{\nu}} = -2 \ln \frac{L\left(\vec{\nu}(\hat{\vec{\theta}})\right)}{L(\hat{\vec{\nu}})} = -2 \sum_{i=1}^{N} \left[n_i \ln \frac{\nu_i(\hat{\vec{\theta}})}{n_i} \right]$$

Wilks 定理 $\boldsymbol{f(t_{\vec{\nu}}|\vec{\nu}) \sim \chi^2(N-1)}$ $\boldsymbol{f(q_{\vec{\nu}}|\vec{\nu}) \sim \chi^2(N-M-1)}$

与泊松数据相比，仅仅少了一个自由度，因为等效地址拟合了$N-1$个参数。