北京大学
PEKING UNIVERSITY

实验物理中的统计方法

第六章：最大似然法

杨振伟

回顾

估计量

估计量好坏的评价标准

相合性
偏倚性
有效性

样本均值

样本方差

样本协方差

相关系数的估计量

不需要概率分布的形式

估计量好坏的三个标准

相合性(一致性)

$$\lim_{n\to\infty}\hat{\theta}=\theta$$

$$\lim_{n\to\infty}P\left(\left|\hat{\theta}-\theta\right|>\varepsilon\right)=0,\forall\varepsilon>0\text{成立}$$

偏倚大小(偏倚性)

$$b=E\left[\hat{\theta}\right]-\theta=0$$

方差大小(有效性)

对任何估计量 $\hat{\theta}'$，都有 $\lim_{n\to\infty}\frac{V\left[\hat{\theta}_n\right]}{V\left[\hat{\theta}_n'\right]}\leq 1$，

则$\hat{\theta}$为渐进有效估计量。

本章要点

- 似然函数，最大似然估计量
- 指数与高斯分布的参数确定
- 最大似然估计量的方差
- 扩展的最大似然估计
- 分区间数据的最大似然估计
- 不等精度观测结果的合并

参数估计与最大似然法

前面给出了在不知道概率密度函数 pdf 的情况下，如果没有未知参数，如何从有限的数据样本中，估计出随机变量的期待值、方差与相关系数。

如果已经知道概率密度函数 pdf 的具体形式，但是函数中包含有未知参数，如何从有限的样本中估计未知参数的期待值、方差与相关系数……

最大似然法

参数估计与概率大小的关系

考虑数据样本$\vec{x}=(x_1,\ldots,x_n)$，x_i独立同分布，$x_i\sim f(x;\theta)$。

目标：估计θ，或更一般地，估计$\vec{\theta}=(\theta_1,\ldots,\theta_m)$

如果$f(x;\theta)$为真，则

$$P(\text{对所有在}[x_i,x_i+\mathrm{d}x_i]\text{观测到的}x_i)=\prod_{i=1}^{n} f(x_i,\theta)\mathrm{d}x_i$$

如果假设(包括θ的取值)为真

→ 可以预料会使观测结果具有高的概率。

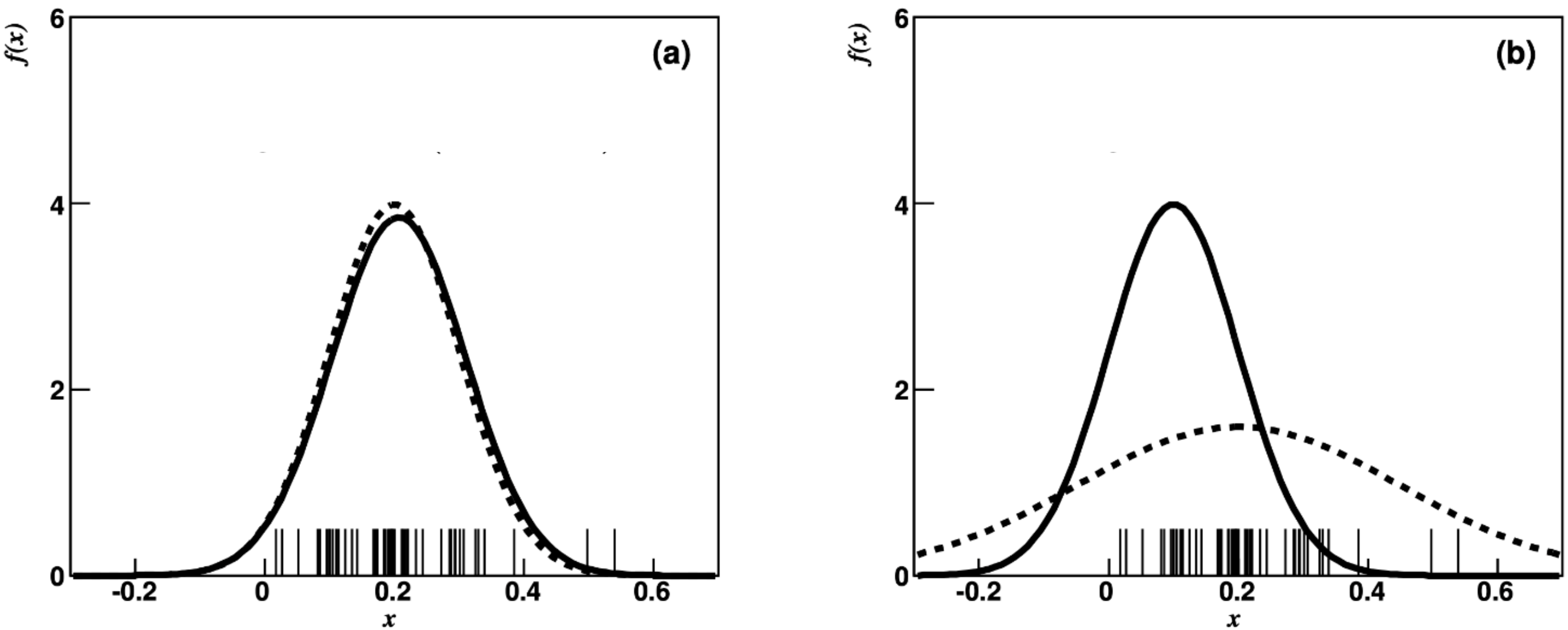
如果假设的θ取值远离真值

→ 会使观测结果具有低的概率。

参数估计与概率大小的关系

$\prod_{i=1}^{n} f(x_i, \theta)$在什么情况下比较大?

$$f(x_i, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$



似然函数

假设一组实验的完整结果可以用矢量$\vec{x} = (x_1, \dots, x_n)$表示，并假设数据$\vec{x}$的联合概率密度依赖于一组参数$\vec{\theta}$，即 $f(\vec{x}; \vec{\theta})$.

利用给定的数据计算$f(\vec{x}; \vec{\theta})$，则$f(\vec{x}; \vec{\theta})$只是 $\vec{\theta}$ 的函数，称为似然函数（likelihood function）。记为

$$L(\vec{\theta}) = f(\vec{x}; \vec{\theta})$$

$\vec{x}$为常数

$\ln L(\vec{\theta})$：对数似然函数

在经典统计中，$L(\theta)$并不是 θ 的概率密度。

在贝叶斯统计中，把$L(\theta) = L(\vec{x}|\theta)$看作给定 θ 的情况下，$\vec{x}$ 的概率密度，然后利用贝叶斯定理得到验后概率密度 $p(\theta|\vec{x})$ 。

独立同分布数据的似然函数

随机变量$x \sim f(x;\vec{\theta})$，考虑x的n次独立观测，观测结果为$x_1, x_2, \dots, x_n$。整个数据样本的联合概率密度为

$$f(x_1, \dots, x_n; \vec{\theta}) = \prod_{i=1}^{n} f(x_i, \vec{\theta})$$

x_i为常数

此时，似然函数为

$$L(\vec{\theta}) = \prod_{i=1}^{n} f(x_i, \vec{\theta})$$

$L(\vec{\theta})$只是 $\vec{\theta}$ 的函数。

对数似然函数：

$$\ln L(\vec{\theta}) = \sum_{i=1}^{n} \ln f(x_i, \vec{\theta})$$

如果假设的$\vec{\theta}$接近于其真实值，则期待得到实际观测数据的概率比较大。

最大似然估计量

定义最大似然估计量 $\hat{\vec{\theta}}$ 为使得 $L(\vec{\theta})$ 最大的 $\vec{\theta}$ 值。

$$\frac{\partial L(\vec{\theta})}{\partial \theta_i} = 0, \qquad i = 1, \dots, m$$

→ $\hat{\vec{\theta}} = (\hat{\theta}_1, \dots, \hat{\theta}_m)$

有时候 $L(\vec{\theta})$可以有好几个极大值 → 取最大值

注意：1）利用了数据的所有信息，与数据分区间无关；
2）定义的最大似然估计量并不保证性质总是最优的。

→ 需要研究诸如无偏性、有效性等问题

实际应用中，最大似然估计量通常能给出最好的实用解。

对数函数是单调函数

$$\frac{\partial L(\vec{\theta})}{\partial \theta_i} = 0 \Leftrightarrow \frac{\partial \ln L(\vec{\theta})}{\partial \theta_i} = 0$$

最大似然估计量的唯一性

考虑 θ 的最大似然估计值是下列方程的解

$$\frac{\partial \ln L(\theta)}{\partial \theta} = 0$$

如果选用另一个等价参数 $h(\theta)$，则 h 的最大似然估计值是下列方程的解

$$\frac{\partial \ln L(\theta)}{\partial h} = 0$$

因为
$$\frac{\partial \ln L(\theta)}{\partial \theta} = \frac{\partial \ln L(\theta)}{\partial h}\frac{\partial h}{\partial \theta}$$
所以，只要 $\frac{\partial h}{\partial \theta} \neq 0$，就有

$$\left.\frac{\partial \ln L(\theta)}{\partial h(\theta)}\right|_{\theta=\hat{\theta}} = \left.\frac{\partial \ln L(\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}} = 0 \Rightarrow \hat{h} = h(\hat{\theta})$$

因此，h的最大似然估计值与参数选取无关，具有唯一性。

本章要点

- 似然函数，最大似然估计量
- 指数与高斯分布的参数确定
- 最大似然估计量的方差
- 扩展的最大似然估计
- 分区间数据的最大似然估计
- 不等精度观测结果的合并

例：指数分布参数的最大似然估计

考虑指数分布

$$f(t;\tau) = \frac{1}{\tau} e^{-\frac{t}{\tau}}$$

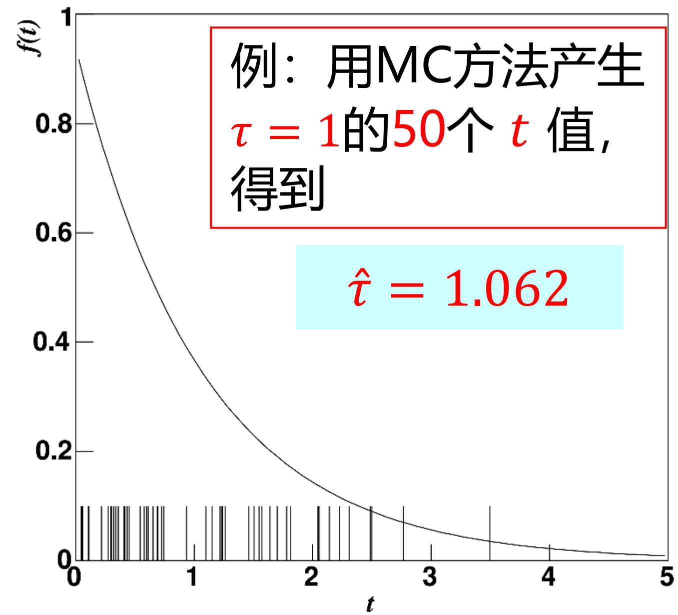
假设简单数据样本 $t_1, \ldots, t_n$。对数似然函数为

$$\ln L(\tau) = \sum_{i=1}^{n} \ln \frac{1}{\tau} e^{-\frac{t_i}{\tau}} = \sum_{i=1}^{n} \left(\ln \frac{1}{\tau} - \frac{t_i}{\tau} \right)$$

令 $\frac{\partial \ln L(\tau)}{\partial \tau} = 0 \;\Rightarrow$ 最大似然估计值

$$\hat{\tau} = \frac{1}{n} \sum_{i=1}^{n} t_i$$

这是是平均寿命的最大似然估计



指数分布中参数最大似然估计的偏倚性

对样本求样本均值

$$\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n} t_i$$

它是 τ 的无偏估计量吗?

原则上可以找出$\hat{\tau}$的概率密度 $g(\hat{\tau};\tau)$，并计算偏倚量大小：（例如，蒙特卡罗方法或特征函数法）

$$b = E[\hat{\tau}] - \tau$$

⇒ 检查估计量是否无偏

但是...

指数分布中参数最大似然估计的偏倚性

一种较简单的方法是直接计算 $E[\hat{\tau}]$：

$$E[\hat{\tau}(t_1,\ldots,t_n)] = \int\cdots\int \hat{\tau}(\vec{t}) f_{\text{joint}}(\vec{t};\tau)\mathrm{d}t_1\ldots\mathrm{d}t_n$$

$$= \int\cdots\int\left(\frac{1}{n}\sum_{i=1}^{n}t_i\right)\frac{1}{\tau}e^{-\frac{t_1}{\tau}}\cdots\frac{1}{\tau}e^{-\frac{t_n}{\tau}}\,\mathrm{d}t_1\ldots\mathrm{d}t_n$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\int t_i\frac{1}{\tau}e^{-\frac{t_i}{\tau}}\mathrm{d}t_i\prod_{j\neq i}\int\frac{1}{\tau}e^{-\frac{t_j}{\tau}}\mathrm{d}t_j\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}\tau = \tau$$

更简单的方法是证明样本均值 $\bar{t}$ 是 均值 $E[t]$ 的无偏估计量，而对于指数分布，$E[t] = \tau$。

因此，$\hat{\tau}$ 是 τ 的无偏估计量。

指数分布中τ的倒数的最大似然估计

假设指数概率密度可写为

$$f(t;\lambda) = \lambda e^{-\lambda t}$$

其中 $\lambda = 1/\tau$，是衰变常数或寿命的倒数。

在测量值是时间 t 的情况下, 如何找出 λ 的最大似然估计量?

根据最大似然估计量的唯一性，当 λ 是 t 的等价参数时,

使 $L_\lambda(\lambda)$ 达到最大值的是 $\lambda(\hat{\tau})$，变量 $\hat{\tau}$ 也同时使 $L_\tau(\tau)$达到最大。

➡ 函数 $\lambda(\theta)$ 的最大似然估计量是 $\hat{\lambda} = \lambda(\hat{\tau})$

$$\hat{\lambda} = \frac{1}{\hat{\tau}} = \left(\frac{1}{n}\sum_{i=1}^{n} t_i\right)^{-1}$$

指数分布中τ的倒数的有偏估计

$\hat{\lambda}$ 是 λ 的无偏估计量吗?

一般而言，一个无偏估计量的非线性函数对该函数的估计是有偏倚的。

利用特征函数方法，可以证明

$$E[\hat{\lambda}] = \lambda \frac{n}{n-1}$$

$$b = E[\hat{\lambda}] - \lambda = \frac{\lambda}{n-1}$$

$\hat{\lambda}$ 是有偏估计量。
当 $n \to \infty$ 时，偏倚 $b \to 0$，称为渐进无偏估计量。

高斯分布中的参数 μ 和 σ^2

考虑高斯变量的简单样本，参数 μ 和 σ^2 未知。其对数似然函数为

$$\ln L(\mu,\sigma^2) = \sum_{i=1}^{n} \ln f(x_i;\mu,\sigma^2)$$
$$= \sum_{i=1}^{n} \left(\ln \frac{1}{\sqrt{2\pi}} + \frac{1}{2} \ln \frac{1}{\sigma^2} - \frac{(x_i-\mu)^2}{2\sigma^2} \right)$$

对 μ 和 σ^2 求导，令导数为零得到

$$\frac{\partial \ln L(\mu,\sigma^2)}{\partial \mu} = \sum_{i=1}^{n} \frac{x_i-\mu}{\sigma^2} = 0 \quad \Rightarrow \quad \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\frac{\partial \ln L(\mu,\sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\sigma^4} = 0 \quad \Rightarrow \quad \widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n} (x_i-\hat{\mu})^2$$

高斯分布参数估计值的偏倚性

前面已证明 $\hat{\mu}$ 是 μ 的无偏估计量。

计算 $\widehat{\sigma^2}$ 的数学期望：

$$E\left[\widehat{\sigma^2}\right] = \frac{n-1}{n}\sigma^2$$

因此，σ^2 的最大似然估计量是有偏估计量。
这种偏倚性随样本容量 n 趋于无穷大而消失。

样本方差

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \hat{\mu})^2$$

对任何概率密度函数的方差估计都是无偏的。

但样本方差不是最大似然估计量。

本章要点

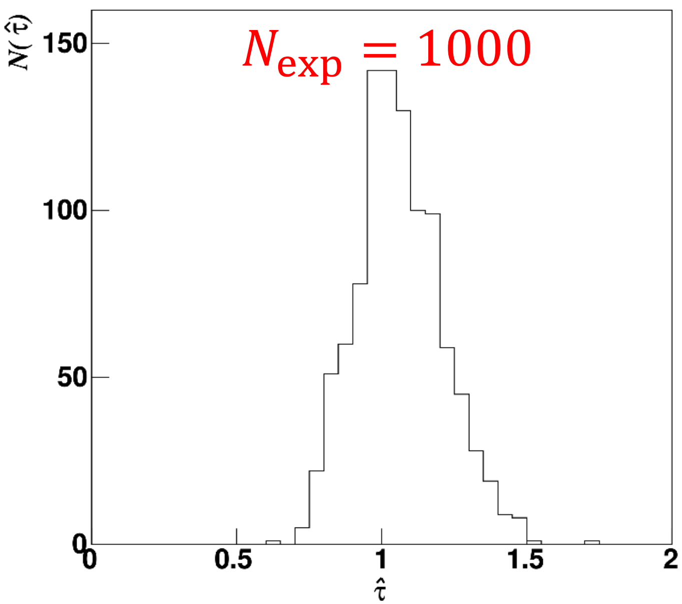
- 似然函数，最大似然估计量
- 指数与高斯分布的参数确定
- 最大似然估计量的方差
- 扩展的最大似然估计
- 分区间数据的最大似然估计
- 不等精度观测结果的合并

估计量的方差：蒙特卡罗方法

我们希望估计并报道参数估计值的统计不确定度，即如果重复整个实验很多次，估计的参数值分布多宽。

一种做法是，利用蒙特卡罗方法多次模拟整个实验，得到 $g(\hat{\theta};\theta,n)$。

例如，对指数分布的例子，$\hat{\tau}=1.062$。以此为 τ 的真值，产生样本容量 $n=50$ 的样本，重复 $N_{\text{exp}}=1000$ 次。每次实验估计 $\hat{\tau}$，并填充直方图。



蒙特卡罗实验可给出样本标准差

$$\hat{\sigma}_{\hat{\tau}}=\left[\frac{1}{N_{\text{exp}}-1}\sum_{i=1}^{N_{\text{exp}}}\left(\hat{\tau}_i-\bar{\hat{\tau}}\right)^2\right]^{1/2}=0.151$$

注意：$g(\hat{\tau};\tau,n)$ 近似服从高斯分布。在大样本极限下，最大似然估计量几乎总是服从高斯分布。

估计量的方差：解析方法

指数分布平均值的估计量为： $\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n} t_i$

对应的概率密度 $g(\hat{\tau};\tau,n)$ 分布的宽度可以估计

$$
\begin{aligned}
V[\hat{\tau}] &= E[\hat{\tau}^2] - (E[\hat{\tau}])^2 \\
&= \int\cdots\int\left(\frac{1}{n}\sum_{i=1}^{n} t_i\right)^2 \frac{1}{\tau}\, e^{-t_1/\tau} \ldots \frac{1}{\tau}\, e^{-t_n/\tau}\, \mathrm{d}t_1\cdots\mathrm{d}t_n \\
&\quad -\left(\int\cdots\int\left(\frac{1}{n}\sum_{i=1}^{n} t_i\right)\frac{1}{\tau}\, e^{-t_1/\tau} \ldots \frac{1}{\tau}\, e^{-t_n/\tau}\, \mathrm{d}t_1\cdots\mathrm{d}t_n\right)^2 \\
&= \frac{\tau^2}{n}
\end{aligned}
$$

$\hat{\tau}$ 的方差是 t 的方差 $1/n$ 倍。

估计量的方差：解析方法(续)

注意：对于未知真值 τ 的 $V[\hat{\tau}]$或 $\sigma_{\hat{\tau}}$，其估计可以采用

$$\hat{\sigma}_{\hat{\tau}} = \frac{\hat{\tau}}{\sqrt{n}} = \frac{1.062}{\sqrt{50}} = 0.150$$

通常情况下，“统计不确定度”由上式给出。例如，

$$\hat{\tau} \pm \hat{\sigma}_{\hat{\tau}} = 1.062 \pm 0.150$$

这就意味着

最大似然法对 τ 的估计值为1.062;

最大似然法对 $g(\hat{\tau};\tau,n)$ 的 σ 的估计为 0.150。

如果 $g(\hat{\tau};\tau,n)$是高斯分布，则$[\hat{\tau}-\hat{\sigma}_{\hat{\tau}}, \hat{\tau}+\hat{\sigma}_{\hat{\tau}}]$为所谓的

“置信水平68.3%的置信区间”

估计量的方差：信息不等式

信息不等式为任何估计量的方差设定了下界

$$V[\hat{\theta}] \geq \boxed{\left(1+\frac{\partial b}{\partial \theta}\right)^2 \Big/ E\left[-\frac{\partial^2 \ln L}{\partial \theta^2}\right]}$$

最小方差界（MVB）

（ b为偏倚量)

也称为 Rao-Cramér-Frechet （RCF）不等式。

如果 $b=0$且等式成立，则称 $\hat{\theta}$ 是有效估计量。

通常偏移量 b 很小，等式严格或在很好的近似下成立，则

$$V[\hat{\theta}] \approx -1 \Big/ E\left[\frac{\partial^2 \ln L}{\partial \theta^2}\right]$$

最大似然估计量对大样本容量 n 几乎总是有效估计量。

通常假设ML估计量为有效估计量，利用RCF边界估计 $V[\hat{\theta}]$。

信息不等式：指数分布的例子

对于指数分布，

$$\ln L(\tau) = \sum_{i=1}^{n}\left(\ln\frac{1}{\tau} - \frac{t_i}{\tau}\right)$$

$$\frac{\partial \ln L(\tau)}{\partial \tau} = \sum_{i=1}^{n}\left(-\frac{1}{\tau} + \frac{t_i}{\tau^2}\right)$$

可以得到

$$\frac{\partial^2 \ln L(\tau)}{\partial \tau^2} = \sum_{i=1}^{n}\left(\frac{1}{\tau^2} - \frac{2t_i}{\tau^2}\right) = \frac{n}{\tau^2}\left(1 - \frac{2}{\tau}\frac{1}{n}\sum_{i=1}^{n} t_i\right) = \frac{n}{\tau^2}\left(1 - \frac{2\hat{\tau}}{\tau}\right)$$

已知 $b = 0$，所以

$$V[\hat{\tau}] \geq \frac{1}{E\left[-\frac{n}{\tau^2}\left(1 - \frac{2\hat{\tau}}{\tau}\right)\right]} = \frac{1}{-\frac{n}{\tau^2}\left(1 - \frac{2E[\hat{\tau}]}{\tau}\right)} = \frac{\tau^2}{n}$$

$$= V[\bar{\tau}]$$

→ ML估计量 $\hat{\tau}$ 对任何样本容量 n 都是有效估计量。

多维参数的信息不等式：费舍尔信息矩阵

对于 m 个参数 $\vec{\theta} = (\theta_1, \ldots, \theta_m)$，最小方差界由费舍尔信息矩阵给出：

$$I_{ij} = E\left[-\frac{\partial^2 \ln L}{\partial\theta_i \partial\theta_j}\right] = -n\int f\left(x;\vec{\theta}\right)\frac{\partial^2 \ln f\left(x;\vec{\theta}\right)}{\partial\theta_i \partial\theta_j}\mathrm{d}x$$

信息不等式：$V - I^{-1}$是半正定矩阵，其中$V_{ij} = \mathrm{cov}\left[\hat{\theta}_i, \hat{\theta}_j\right]$。

因此，
$$V\left[\hat{\theta}_i\right] \geq (I^{-1})_{ii}$$

经常用 I^{-1} 近似协方差矩阵，利用 L 在最大值处的二阶微分矩阵来估计。

$$(V^{-1})_{ij} = -\left.\frac{\partial^2 \ln L}{\partial\theta_i \partial\theta_j}\right|_{\vec{\theta}=\hat{\vec{\theta}}}$$

费舍尔信息矩阵

一维：
$$\hat{V}(\hat{\theta}) = \left.\left(-\frac{\partial^2 \ln L}{\partial\theta^2}\right)^{-1}\right|_{\theta=\hat{\theta}}$$

估计量的方差: 图解法

考虑单参数 θ 的情况，将 $\ln L(\theta)$ 在 $\hat{\theta}$ 附近展开：

$$\ln L(\theta) = \underbrace{\boxed{\ln L(\hat{\theta})}}_{=\ln L_{\max}} + \underbrace{\boxed{\left[\frac{\partial \ln L}{\partial \theta}\right]_{\theta=\hat{\theta}} (\theta - \hat{\theta})}}_{=0} + \frac{1}{2!}\left[\frac{\partial^2 \ln L}{\partial \theta^2}\right]_{\theta=\hat{\theta}} (\theta - \hat{\theta})^2 + \cdots$$

信息不等式 ➔

$$\left[\frac{\partial^2 \ln L}{\partial \theta^2}\right]_{\theta=\hat{\theta}} = \frac{1}{\widehat{\sigma^2}_{\hat{\theta}}}$$

$$\log L(\theta) \approx \log L_{\max} - \frac{(\theta - \hat{\theta})^2}{2\widehat{\sigma^2}_{\hat{\theta}}}$$

即

$$\log L(\hat{\theta} \pm \hat{\sigma}_{\hat{\theta}}) \approx \log L_{\max} - \frac{1}{2}$$

为了得到$\hat{\sigma}_{\hat{\theta}}$，让$\theta$偏离$\hat{\theta}$，使得$\ln L$减小数值 0.5。

估计量的方差: 图解法示例

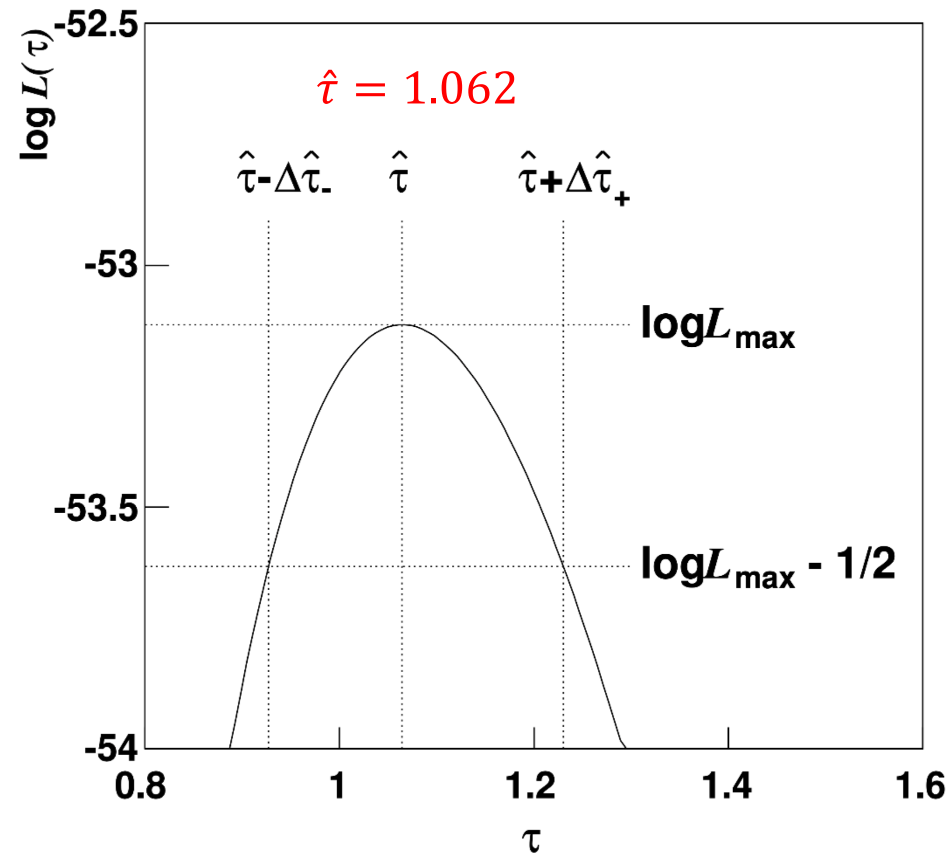
考虑指数分布最大似然估计的例子:

$$\hat{\tau} = 1.062$$

$$\Delta\hat{\tau}_- = 0.137$$

$$\Delta\hat{\tau}_+ = 0.165$$

$$\hat{\sigma}_{\hat{\tau}} \approx \Delta\hat{\tau}_- \approx \Delta\hat{\tau}_+ \approx 0.15$$

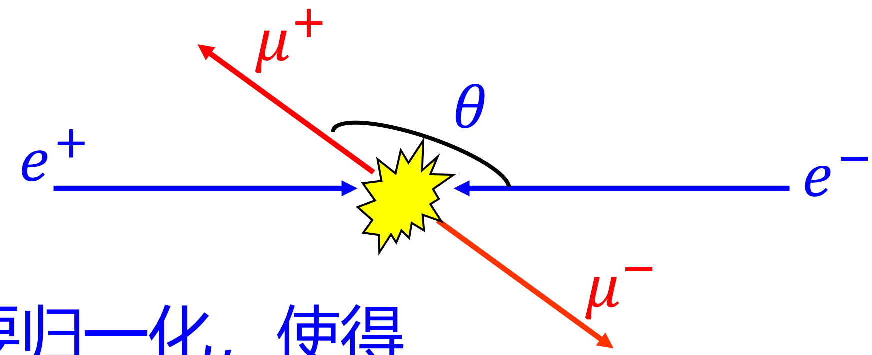


$\ln L\,(\tau)$不是抛物线，因为样本容量 $n = 50$ 不足够大。

双参数的最大似然估计

考虑某个散射角的分布，定义 $x = \cos\theta$

$$f(x;\alpha,\beta) = \frac{1+\alpha x+\beta x^2}{2+2\beta/3}$$



如果 $x_{\min} < x < x_{\max}$，则需要归一化，使得

$$\int_{x_{\min}}^{x_{\max}} f(x;\alpha,\beta)\mathrm{d}x = 1$$

例：$\alpha = 0.5, \beta = 0.5, x_{\min} = -0.95, x_{\max} = 0.95$

产生 $n = 2000$ 个蒙特卡罗事例。

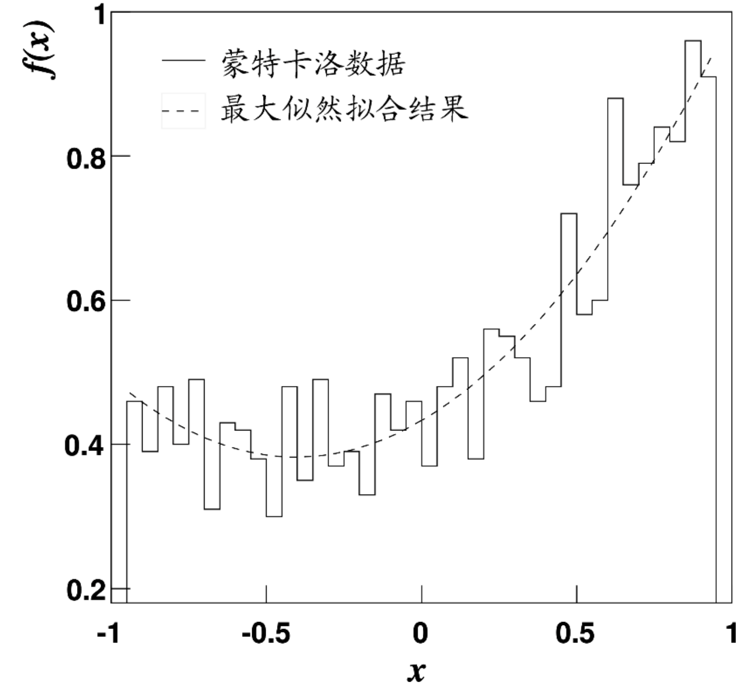
似然函数：

$$L(\alpha,\beta) = \prod_{i=1}^{2000} f(x_i;\alpha,\beta)$$

双参数的最大似然估计（续）

数值求 $\ln L(\alpha, \beta)$ 的最大值，得到

$$\hat{\alpha} = 0.502$$
$$\hat{\beta} = 0.509$$



协方差矩阵

$$\left(\hat{V}^{-1}\right)_{ij} = -\left.\frac{\partial^2 \log L}{\partial\theta_i \partial\theta_j}\right|_{\vec{\theta}=\hat{\vec{\theta}}}$$

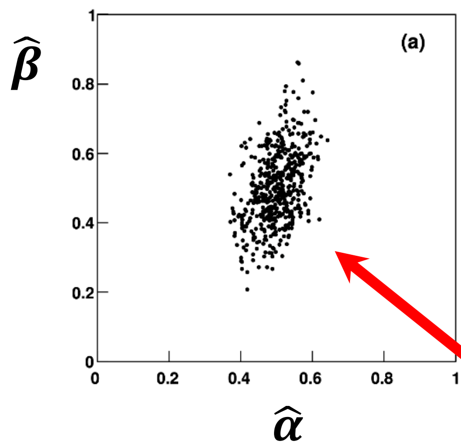
$$\hat{\alpha} = 0.50 \pm 0.05$$
$$\hat{\beta} = 0.51 \pm 0.11$$
$$\widehat{\mathrm{cov}}[\hat{\alpha}, \hat{\beta}] = 0.0024$$
$$\hat{\rho} = r = 0.42$$

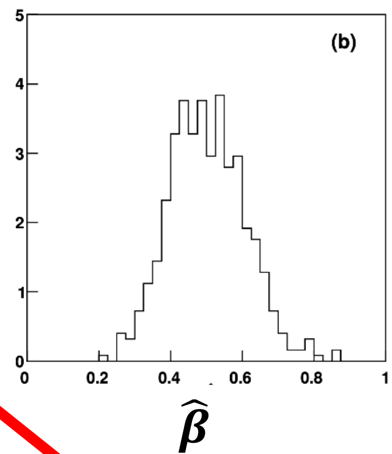
注意：拟合中采用的是点拟合，与直方图的区间大小划分无关。

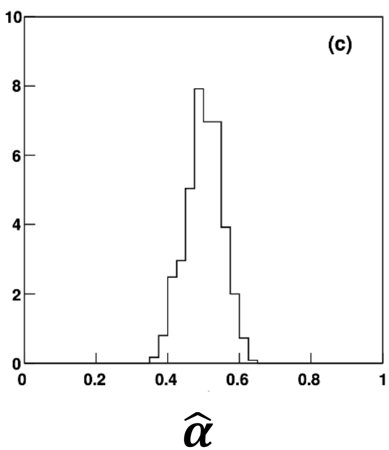
但是，可以与直方图比较得到拟合优度。

双参数拟合结果的蒙特卡罗检验

做500次模拟实验，每次模拟样本容量均为$n = 2000$。重复最大似然拟合。







表明 $\hat{\alpha}$ 和 $\hat{\beta}$ 不独立

$$\overline{\hat{\alpha}} = 0.499 \qquad \overline{\hat{\beta}} = 0.498$$

$$s_{\hat{\alpha}} = 0.051 \qquad s_{\hat{\alpha}} = 0.111$$

$$\widehat{\mathrm{cov}}\left[\hat{\alpha}, \hat{\beta}\right] = 0.0024$$

$$\hat{\rho} = r = 0.42$$

$\hat{\alpha}$ 和 $\hat{\beta}$ 正相关

边缘概率密度函数近似为高斯分布。

双参数拟合中的$(\ln L_{\max} - 1/2)$等高线

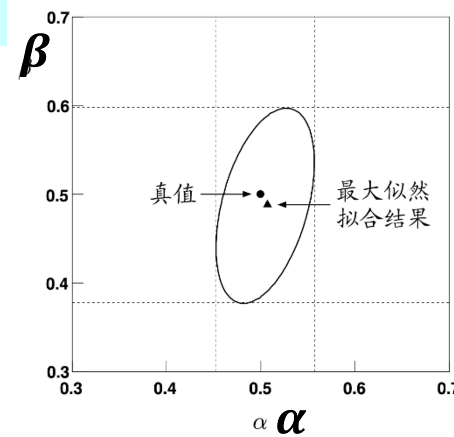
对于大样本容量 n，$\ln L$ 在最大值处具有二次型的形式：

$$\ln L(\alpha,\beta) \approx \ln L_{\max} - \frac{1}{2(1-\rho^2)}\left[\left(\frac{\alpha-\hat{\alpha}}{\sigma_{\hat{\alpha}}}\right)^2 + \left(\frac{\beta-\hat{\beta}}{\sigma_{\hat{\beta}}}\right)^2 - 2\rho\left(\frac{\alpha-\hat{\alpha}}{\sigma_{\hat{\alpha}}}\right)\left(\frac{\beta-\hat{\beta}}{\sigma_{\hat{\beta}}}\right)\right]$$

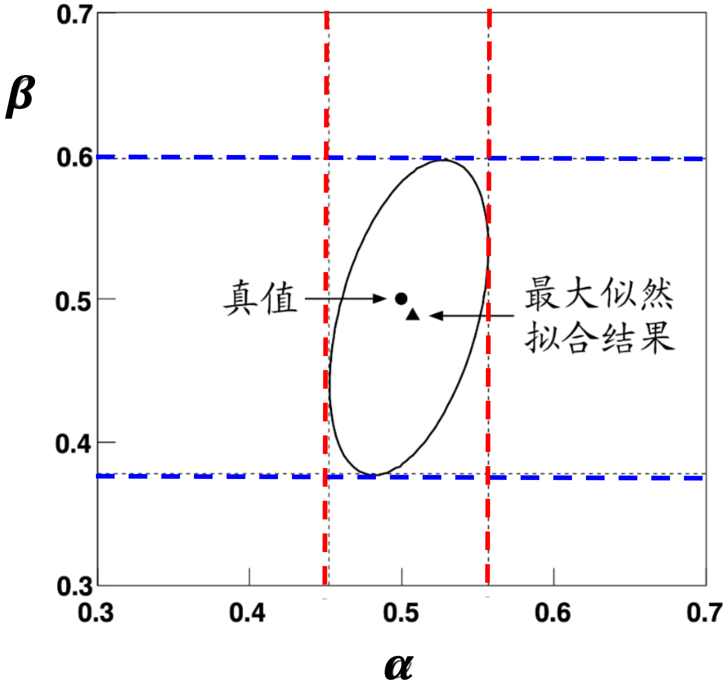
等高线 $\ln L(\alpha,\beta) = \ln L_{\max} - \frac{1}{2}$ 是个椭圆：

$$\frac{1}{(1-\rho^2)}\left[\left(\frac{\alpha-\hat{\alpha}}{\sigma_{\hat{\alpha}}}\right)^2 + \left(\frac{\beta-\hat{\beta}}{\sigma_{\hat{\beta}}}\right)^2 - 2\rho\left(\frac{\alpha-\hat{\alpha}}{\sigma_{\hat{\alpha}}}\right)\left(\frac{\beta-\hat{\beta}}{\sigma_{\hat{\beta}}}\right)\right] = 1$$

在蒙特卡罗样本中，每次的拟合结果对应于 $\beta-\alpha$平面上的一个点。



从 $\ln L$ 等高线估计方差、协方差



$\beta-\alpha$平面：

椭圆是 $\ln L(\alpha,\beta)=\ln L_{\max}-\frac{1}{2}$ 的等高线

椭圆的切线给出参数估计的标准差

椭圆的倾角ϕ与相关性有关：

$$\tan 2\phi=\frac{2\rho\sigma_{\hat{\alpha}}\sigma_{\hat{\beta}}}{\sigma_{\hat{\alpha}}^2-\sigma_{\hat{\beta}}^2}$$

注意：估计量之间的相关性大小会影响估计量的标准差（统计不确定度），或偏大或偏小。

本章要点

- 似然函数，最大似然估计量
- 指数与高斯分布的参数确定
- 最大似然估计量的方差
- 扩展的最大似然估计
- 分区间数据的最大似然估计
- 不等精度观测结果的合并

扩展的最大似然法

前面只考虑了样本容量 n 固定的情形，有时候，n 被看作均值为 ν 的泊松变量。

➡ 实验结果定义为：$n, x_1, \dots, x_n$

则扩展的似然函数为

$$L(\nu, \vec{\theta}) = \frac{\nu^n}{n!} e^{-\nu} \prod_{i=1}^{n} f(x_i; \vec{\theta})$$

假设理论给出 $\nu = \nu(\vec{\theta})$，去掉与 $\vec{\theta}$ 无关的项，对数似然函数：

$$\ln L(\vec{\theta}) = n \ln \nu(\vec{\theta}) - \nu(\vec{\theta}) + \sum_{i=1}^{n} \ln\left(f(x_i; \vec{\theta})\right)$$

$$= -\nu(\vec{\theta}) + \sum_{i=1}^{n} \ln\left(\nu(\vec{\theta}) \cdot f(x_i; \vec{\theta})\right)$$

扩展的最大似然法（续）

例：期待事例数 $\nu(\vec{\theta}) = \sigma(\vec{\theta}) \int L dt$,

其中总截面$\sigma(\vec{\theta})$预期为理论参数的函数。

扩展的最大似然法利用了更多信息 ➔ $\hat{\vec{\theta}}$的不确定度更小

这对某些测量可能会非常重要。

扩展的最大似然法：ν和$\vec{\theta}$独立

假如 ν 和 $\vec{\theta}$ 相互独立，

$$L(\nu,\vec{\theta}) = \frac{\nu^n}{n!}e^{-\nu}\prod_{i=1}^{n} f(x_i;\vec{\theta})$$

$$\frac{\partial L}{\partial \nu} = \left(\frac{n}{\nu} - 1\right)\frac{\nu^n}{n!}\,e^{-\nu}\prod_{i=1}^{n} f(x_i;\vec{\theta}) = 0 \quad \Rightarrow \quad \hat{\nu} = n$$

$$\frac{\partial L}{\partial \vec{\theta}} = 0 \quad \Rightarrow \quad \text{最大似然估计}$$

适用于处理 $f(x;\vec{\theta})$ 已知的不同成分叠加的情况。

例如，信号成分和本底成分的叠加。

可将其分解成单独求 ν 和 $\vec{\theta}$ 估计值的问题。

扩展的最大似然法：ν和$\vec{\theta}$独立（续）

如果联合概率密度函数可以表示为

$$f(x;\vec{\theta}) = \sum_{i=1}^{m} \theta_i f_i(x)$$

根据概率的定义可知并非所有的 θ_i 独立：

$$\sum_{i=1}^{m} \theta_i = 1 \Rightarrow \theta_m f_m(x) = \left(1 - \sum_{i=1}^{m-1} \theta_i\right) f_m(x)$$

在扩展的最大似然法中

$$\ln L(\nu,\vec{\theta}) = -\nu + \sum_{i=1}^{n} \ln\left(\sum_{j=1}^{m} \nu\theta_j f_j(x_i)\right)$$

定义 $\mu_i = \nu\theta_i$

$$\ln L(\vec{\mu}) = -\sum_{j=1}^{m} \mu_j + \sum_{i=1}^{n} \ln\left(\sum_{j=1}^{m} \mu_j f_j(x_i)\right)$$

μ_j为类型 j 的事例数期待值。n 为观测事例总数。

扩展的最大似然法：非物理结果

假设有两类事例：信号 (s) 与本底 (b)

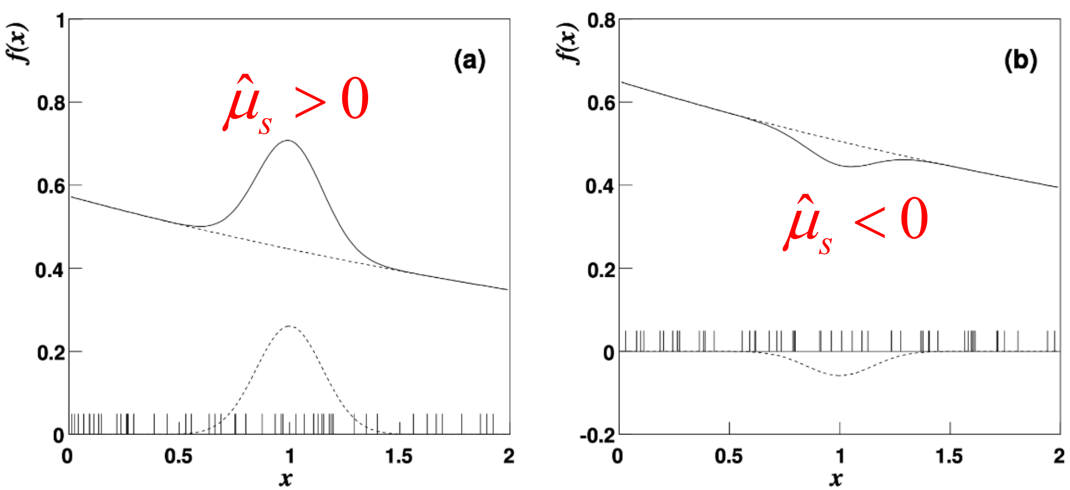
$$f(x;\mu_s,\mu_b)=\frac{\mu_s}{\mu_s+\mu_b}f_s(x)+\frac{\mu_b}{\mu_s+\mu_b}f_b(x)$$

假设 $f_s(x)$ 和 $f_b(x)$ 已知，需要估计 μ_s 和 μ_b

$$\ln L(x;\mu_s,\mu_b)=-(\mu_s+\mu_b)+\sum_{i=1}^{n}\ln[(\mu_s+\mu_b)f(x_i;\mu_s,\mu_b)]$$

本底高低对拟合结果不确定度影响很大。

问题：如果出现负值，应该如何报告结果？



本章要点

- 似然函数，最大似然估计量
- 指数与高斯分布的参数确定
- 最大似然估计量的方差
- 扩展的最大似然估计
- 分区间数据的最大似然估计
- 不等精度观测结果的合并

分区间的最大似然估计

通常数据 $\vec{x}$ 在划分为 N 个区间的直方图中的频数为

$$\vec{n} = (n_1, \dots, n_N), \qquad n_{\text{tot}} = \sum_{i=1}^{N} n_i$$

通常称“对直方图拟合”

在某种假设下，频数期待值为

$$\nu_i(\vec{\theta}) = \nu_{\text{tot}} \int_{x_i^{min}}^{x_i^{\max}} f(x; \vec{\theta})\, \mathrm{d}x$$

$$\vec{\nu} = (\nu_1, \dots, \nu_N), \qquad \nu_{\text{tot}} = \sum_{i=1}^{N} \nu_i$$

如果用多项分布描述样本（ n_{tot}为常数 ）:

$$f(\vec{n}; \vec{\nu}) = \frac{n_{\text{tot}}!}{n_1! \cdots n_N!} \left(\frac{\nu_1}{n_{\text{tot}}}\right)^{n_1} \cdots \left(\frac{\nu_N}{n_{\text{tot}}}\right)^{n_N}$$

$$\ln L(\vec{\theta}) = \sum_{i=1}^{N} n_i \ln \nu_i(\vec{\theta})$$

分区间的最大似然估计：指数分布

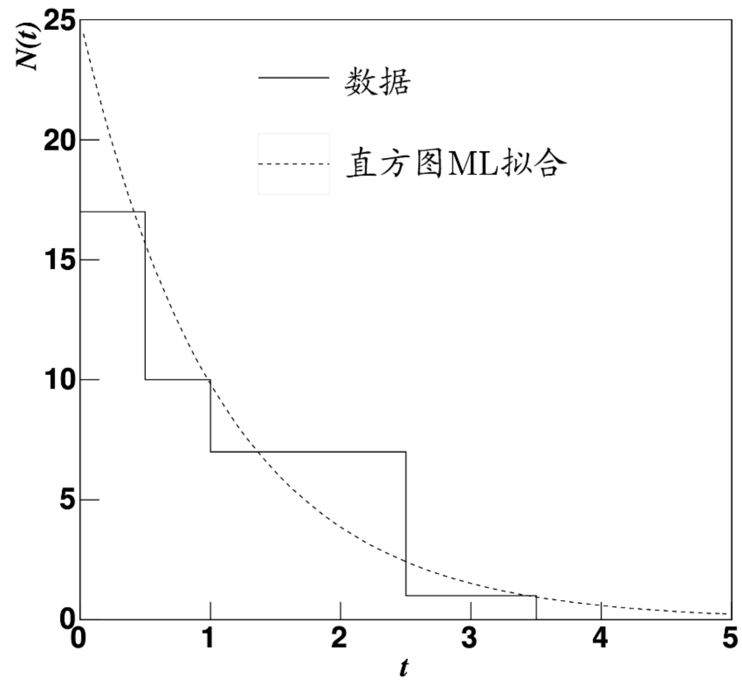
把指数分布的样本填充成直方图

直方图拟合（binned ML）：

$$\hat{\tau} = 1.07 \pm 0.17$$

点估计（unbinned ML）：

$$\hat{\tau} = 1.06 \pm 0.15$$



两者结果吻合。
点估计结果误差较小。

当区间宽度为零时，

直方图拟合 → 点估计

如果 n_i 是泊松随机变量，扩展的对数似然函数为

$$\ln L(\nu_{\text{tot}}, \vec{\theta}) = -\nu_{\text{tot}} + \sum_{i=1}^{N} n_i \ln \nu_i(\nu_{\text{tot}}, \vec{\theta})$$

分区间处理数据的问题

分区间处理数据有时会因区间宽度过大而造成部分信息丢失，影响到参数的估计。例如

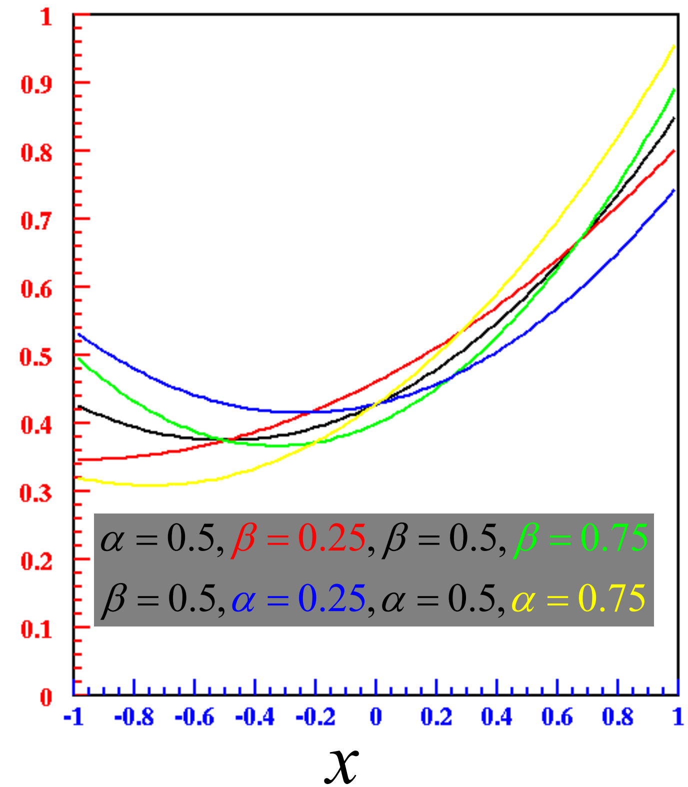
$$f(x;\alpha,\beta)=\frac{1+\alpha x+\beta x^2}{2+2\beta/3}$$

在$x\in[-1,1]$范围内分50个区间

$$\hat{\alpha}=0.47\pm0.05,$$
$$\hat{\beta}=0.39\pm0.11$$

在$x\in[-1,1]$范围内分200个区间

$$\hat{\alpha}=0.50\pm0.05,$$
$$\hat{\beta}=0.50\pm0.11$$



因此，对直方图拟合，一定要确认区间的大小对结果无明显影响。
注意：区间无穷小时，与点估计结果一致。

本章要点

- 似然函数，最大似然估计量
- 指数与高斯分布的参数确定
- 最大似然估计量的方差
- 扩展的最大似然估计
- 分区间数据的最大似然估计
- 不等精度观测结果的合并

不等精度观测结果的合并

对某固定量 μ 作 n 次独立的不等精度测量，结果为 $x_i \pm \sigma_i$，方差已知，且 $x_i \sim N\left(\mu, \sigma_i^2\right)$。则该样本的似然函数为

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

取对数并解似然方程

$$\frac{\partial \ln L(\mu)}{\partial \mu} = \sum_{i=1}^{n} \frac{x_i-\mu}{\sigma_i} = 0 \Rightarrow \hat{\mu} = \frac{\sum_{i=1}^{n} \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}} = \frac{1}{\omega} \sum_{i=1}^{n} \omega_i x_i$$

权重因子

$$\omega_i = 1/\sigma_i^2$$

$$\omega = \sum_{i=1}^{n} \omega_i$$

$$\widehat{\sigma^2}_{\hat{\mu}} = \left({}^{-1}\Big/\frac{\partial^2 \ln L}{\partial \mu^2}\right)\Bigg|_{\mu=\hat{\mu}} = \frac{1}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}} = \frac{1}{\omega}$$

不等精度观测结果的合并：另一种权重

对某一固定量 μ 作 n 次不等精度测量，测量值为 $x_1,\ldots,x_n$，对应的标准差分别为$\sigma_1,\ldots,\sigma_n$ 。每个测量值 $x_i \sim N\left(\mu,\sigma_i^2\right)$，且各测量值相互独立，方差已知。则该样本的似然函数为

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

取对数并解似然方程

$$\frac{\partial\ln L(\mu)}{\partial\mu}=\sum_{i=1}^{n}\frac{x_i-\mu}{\sigma_i}=0 \Rightarrow \hat{\mu}=\frac{\sum_{i=1}^{n}\frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{n}\frac{1}{\sigma_i^2}}=\sum_{i=1}^{n}\omega_i x_i$$

权重因子

$$\omega_i=\frac{1/\sigma_i^2}{\sum_{i=1}^{n}1/\sigma_i^2}$$

$$\widehat{\sigma^2}_{\hat{\mu}}=\left(-1\Big/\frac{\partial^2\ln L}{\partial\mu^2}\right)\Bigg|_{\mu=\hat{\mu}}=\frac{1}{\sum_{i=1}^{n}\frac{1}{\sigma_i^2}}$$

$$\sum_{i=1}^{n}\omega_i=1$$

例：μ轻子寿命的世界平均值

共有五个实验精确测量了 μ 轻子的平均寿命 (单位：ns)

τ_i	权重因子ω_i	$\tau_i - \hat{\tau}$	$\chi_i^2 \equiv \omega_i(\tau_i - \hat{\tau})^2$
2197.078 ± 0.073	187.65	0.045	0.379
2197.025 ± 0.155	41.62	-0.008	0.0027
2196.95 ± 0.06	277.78	-0.083	1.917
2197.11 ± 0.08	156.25	0.077	0.924
2197.3 ± 0.3	11.11	0.267	0.792

$\omega_i = 1/\sigma_i^2$

$\omega = \sum_{i=1}^{5} \omega_i = 674.415\ \mathrm{ns}^2$

$$\hat{\tau} = \frac{1}{\omega}\sum_{i=1}^{n} \omega_i \tau_i = 2197.03\ \mathrm{ns}$$

$$\hat{\sigma} = \frac{1}{\sqrt{\omega}} = 0.04\ \mathrm{ns}$$

$$\sum_{i=1}^{n} \chi_i^2/(5-1) = 4.015/4 = 1.004$$

$\Rightarrow$ 与期望值相符。

总结

- ❑ 最大似然估计量
- ❑ 四种方法给出最大似然估计的方差
  - ✓ 解析方法
  - ✓ 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