北京大学
PEKING UNIVERSITY

实验物理中的统计方法

第五章：参数估计的一般概念

杨振伟

回顾

假设，检验统计量，显著性水平，功效

纽曼-皮尔逊引理

线性检验统计量，费舍尔甄别函数

非线性检验统计量，神经网络

检验拟合优度，p值

信号观测的显著程度

皮尔逊的χ^2检验

本章要点

- 估计量
- 样本均值
- 样本方差
- 样本协方差

再论统计分析的目标

既包含检验假设，
也包含对假设中可能存在的参数进行估计。

实际问题往往是：

有限样本
参数本身不是直接观测量

如何给出参数的最佳估计(包括均值、方差和协方差等)

1. 如何构造参数估计量问题
2. 对参数估计量的评估问题
3. 对参数的有效估计问题

参数估计

概率密度函数的参数是表征其形状的常数，例如

$$f(x;\theta)=\frac{1}{\theta}e^{-\frac{x}{\theta}}$$

随机变量　　参数

假设对随机变量x的n次独立观测得到样本$\vec{x}=(x_1,\dots,x_n)$

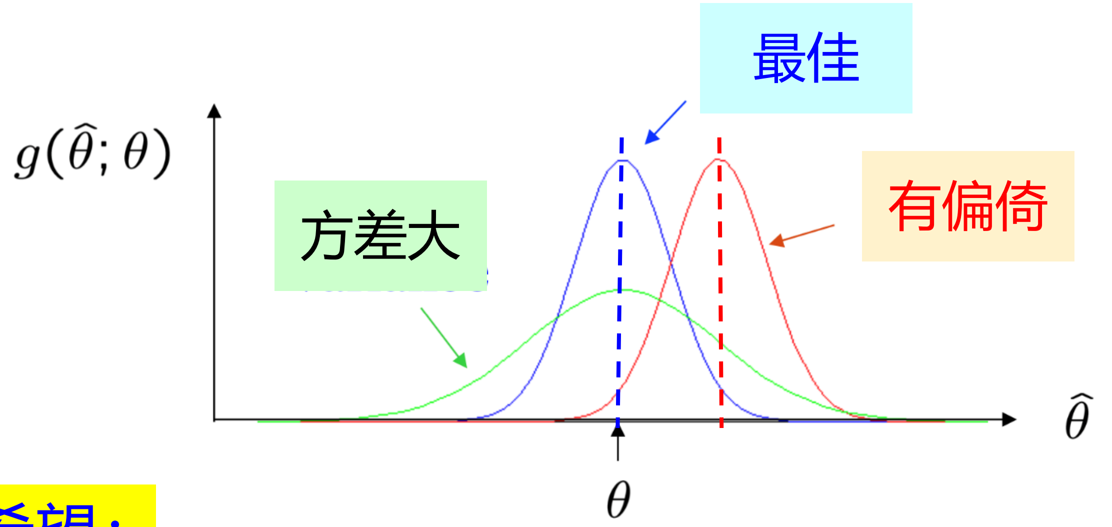
我们希望找到数据样本的适当函数，用来估计参数θ：

估计量：　$\hat{\theta}(\vec{x})$　　估计量用符号“$\hat{}$”表示

估计量是数据样本的函数；对于给定数据样本，估计量的结果称为估计值

估计量的性质

重复整个测量，每次得到的估计值将服从某个分布$g(\hat{\theta};\theta)$



我们希望：

偏倚为零或很小（系统不确定度）：$b = E[\hat{\theta}] - \theta$

➔ 多次重复测量的均值应当趋于真值

方差小（统计不确定度小）：$V[\hat{\theta}]$

➔ 偏倚小和方差小通常是相互矛盾的要求

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 (无偏性)

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

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。

参数估计量的构造与收敛性

如何构造参数θ的估计量$\hat{\theta}(\vec{x})$?

没有一个完美的办法

首先是要求相合性（consistency）

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

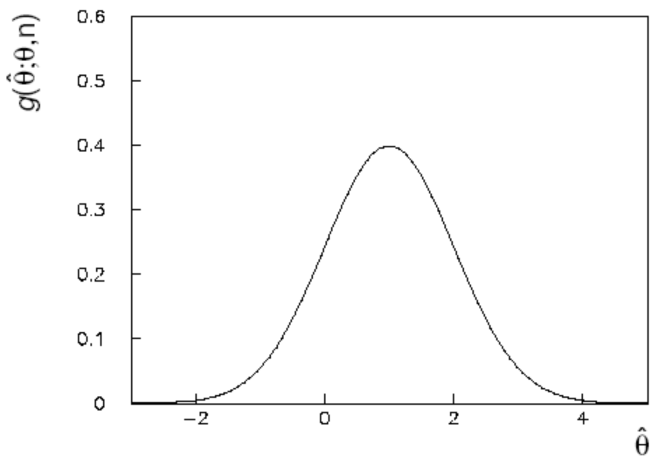
例如，随着样本容量的增大，估计值收敛于真值：

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

注意：在统计意义上的收敛并不保证不会有个别特殊的$\hat{\theta}_{\text{obs}}$与$\theta$真值有较大偏离

估计量的偏倚问题

考虑样本容量n固定的估计量$\hat{\theta}$，其概率密度函数$g(\hat{\theta};\theta,n)$



我们不知道θ真值，只能得到$\hat{\theta}_{\text{obs}}$值

$g(\hat{\theta};\theta,n)$的属性包括：

方差$V[\hat{\theta}]=\sigma_{\hat{\theta}}^2$ ($\sigma_{\hat{\theta}}$：统计不确定度)

偏倚$b=E[\hat{\theta}]-\theta$ (系统不确定度)

对于大统计量 ➡ $\sigma_{\hat{\theta}} \propto \frac{1}{\sqrt{n}}, b \propto \frac{1}{n}$

均方差

为了衡量估计量的好坏，有时会考虑其与真值 θ 的均方误差（mean square error, MSE）

$$
\begin{aligned}
\mathrm{MSE} &= E\left[(\hat{\theta}-\theta)^2\right] = E[\hat{\theta}^2] - 2\theta E[\hat{\theta}] + \theta^2 \\
&= E[\hat{\theta}^2] - 2E\left[\hat{\theta}E[\hat{\theta}]\right] + (E[\hat{\theta}])^2 + (E[\hat{\theta}])^2 - 2\theta E[\hat{\theta}] + \theta^2 \\
&= E\left[(\hat{\theta}-E[\hat{\theta}])^2\right] + (E[\hat{\theta}-\theta])^2 \\
&= V[\hat{\theta}] + b^2
\end{aligned}
$$

➡ 通常要求无偏估计量中，对应的方差达到最小。

"有效估计量"：无偏估计量中方差最小的估计量

有时需要在方差与偏倚之间存在平衡点

无参数的样本均值与样本方差估计

考虑对随机变量x作n次独立测量

➡ 样本大小为 n

等效为对n维矢量$\vec{x} = (x_1, \ldots, x_n)$作单次独立实验

由于x_i相互独立，所以样本的联合概率密度函数可表示为：

$$f_{\mathrm{sample}}(\vec{x}) = f(x_1)f(x_2)\cdots f(x_n)$$

任务：从数据样本推断 $f(x)$ 的属性

➡ 构造数据样本的函数，以便估计 $f(x)$ 的各种属性，包括均值，方差，等等

含参数情况下如何估计参数

通常先给出 $f(x)$ 的假设形式，其中包含未知参数 θ

→ 利用给定的 $f(x,\theta)$ 形式和数据样本估计参数 θ

统计量(statistic) ：数据样本的函数（不含未知参数）
估计量(estimator)：用来估计pdf某些属性的统计量
记号：θ的估计量为$\hat{\theta}$
估计值(estimate) ：估计量的观测值，通常记为$\hat{\theta}_{\text{obs}}$

参数拟合：利用 x 的数据样本估计参数θ的过程

每次实验给出的估计量$\hat{\theta}$是满足概率密度$g(\hat{\theta};\theta)$的随机变量。
估计量$\hat{\theta}$的均值为

$$E[\hat{\theta}(\vec{x})] = \int \hat{\theta}\, g(\hat{\theta};\theta)\mathrm{d}\hat{\theta} = \int \cdots \int \hat{\theta}(\vec{x}) f(x_1;\theta)\cdots f(x_n;\theta)\mathrm{d}x_1\cdots\mathrm{d}x_n$$

均值的估计量与弱大数定理

考虑物理量 x 的 n 个测量 $x_1,\dots,x_n$，我们不知道对应的pdf，想构造一个x_i的函数来估计x的均值，一个可能为

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

(样本均值)

如果$V[x]$有限，$\overline{x}$ 则是一个与 μ 相合的估计量，即

$$\forall\varepsilon>0,\qquad \lim_{n\to\infty}P\left(\left|\frac{1}{n}\sum_{i=1}^{n}x_i-\mu\right|\geq\varepsilon\right)=0$$

弱大数定理

计算期待值

$$E[\overline{x}]=E\left[\frac{1}{n}\sum_{i=1}^{n}x_i\right]=\frac{1}{n}\sum_{i=1}^{n}E[x_i]=\frac{1}{n}\sum_{i=1}^{n}\mu=\mu$$

➡ $\overline{x}$是μ的无偏估计量。

问题：$\overline{x}$的方差是多少?

样本均值的评估量（方差）

对样本均值可靠程度的评估可以用样本均值的方差来估计，

$$
\begin{aligned}
V[\overline{x}] &= E[(E[\overline{x}]-\overline{x})^2] = E\left[\overline{x}^2\right]-(E[\overline{x}])^2 \\
&= E\left[\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)\left(\frac{1}{n}\sum_{j=1}^{n}x_j\right)\right]-\mu^2 \\
&= \frac{1}{n^2}\sum_{i,j=1}^{n}E\left[x_i x_j\right]-\mu^2 \\
&= \frac{1}{n^2}[(n^2-n)\mu^2+n(\mu^2+\sigma^2)]-\mu^2 \\
&= \frac{\sigma^2}{n}
\end{aligned}
$$

这里σ^2是x的方差，并利用了$i \neq j$时

$E\left[x_i x_j\right] = E[x_i]E\left[x_j\right] = \mu^2$　　　　$E\left[x_i^2\right] = \mu^2+\sigma^2$

例：均值的测量精度

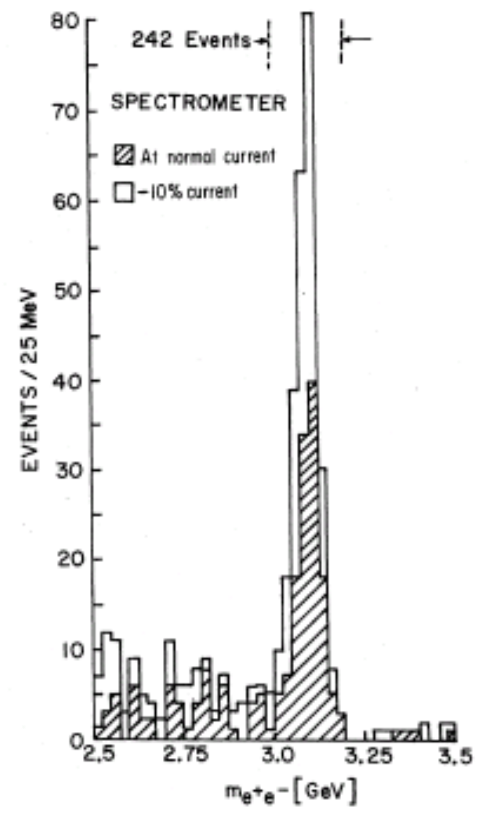
丁肇中在发现 J/ψ 粒子的实验中观测到 25 个 $J/\psi \rightarrow e^+e^-$事例，装置的质量测量精度 $\Delta m/m = 1\%$，质量分布的平均质量是 3.1 GeV。

装置的质量测量精度

$$\sigma_m = 3100\times 1\% = 31\ \mathrm{MeV}$$

如果测量的质量分布与装置的测量精度相同，可以得到

$$\sigma_m = 3100\times 1\%\times \frac{1}{\sqrt{25}} = 6.2\ \mathrm{MeV}$$

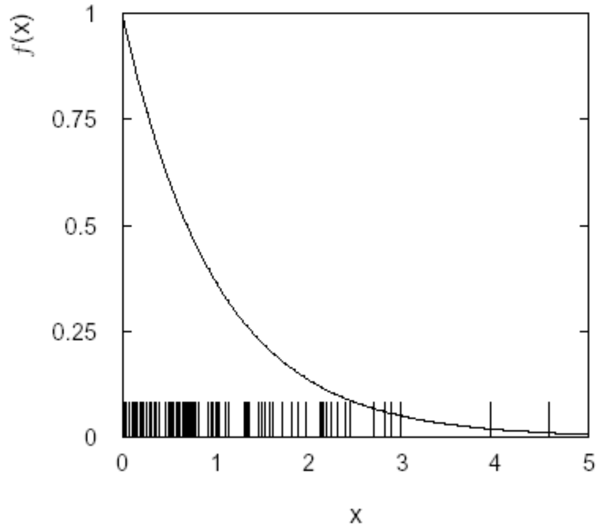


Phys. Rev. Lett. 33, 1404 (1974)

问题：结果应该报告为

3100 ± 31 MeV 还是 3100 ± 6 MeV?

例：均值的估计量与评估量

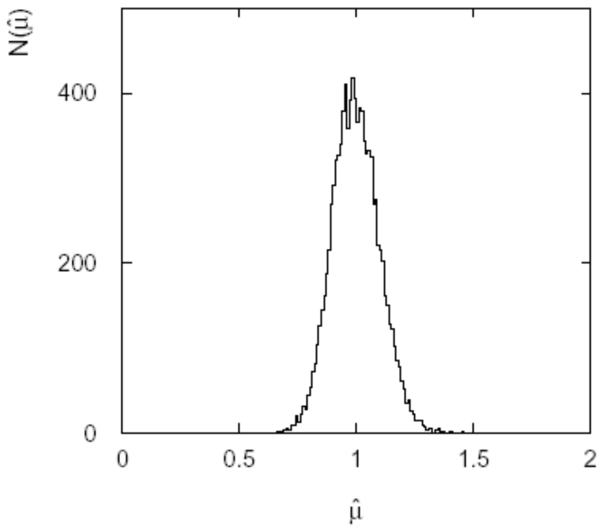


指数分布的蒙特卡罗样本 $n = 100$，寿命真实值为 $\mu = 1$。

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 1.073$$

重复10^4次，每次样本容量都是$n = 100$，把每个样本的样本均值填入直方图

根据中心极限定理，$\hat{\mu}$近似服从高斯分布



$\overline{\hat{\mu}} = 0.9981$ ($\hat{\mu}$无偏)

$\hat{\mu}$的样本标准差为 $0.0995 \approx \sigma/\sqrt{n}$

方差的估计量

假如均值μ和方差 $V[x]=\sigma^2$ 都是未知量，样本方差定义为

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2=\frac{n}{n-1}\left(\overline{x^2}-\overline{x}^2\right)$$

因子 $\frac{1}{n-1}$ 保证 s^2无偏，即 $E[s^2]=\sigma^2$。

假如 $\mu=E[x]$ 先验已知（例如某种假设的预期值），则

$$S^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2=\overline{x^2}-\overline{x}^2$$

s^2和S^2都是σ^2的无偏估计量

s^2 的方差

$$V[s^2]=\frac{1}{n}\left(\mu_4-\frac{n-3}{n-1}\mu_2^2\right)$$

可以利用下式可估计 μ_k

$$m_k=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^k$$

这里μ_k是第k阶中心矩，例如，$\mu_2=\sigma^2$

协方差与相关系数的估计量

协方差 $V_{xy} = \mathrm{cov}[x, y]$ 的估计量为

$$\hat{V}_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \frac{n}{n-1}\left(\overline{xy} - \overline{x}\cdot\overline{y}\right)$$

(无偏)

相关系数 $\rho = V_{xy}/\left(\sigma_x \sigma_y\right)$的估计量为

$$\hat{\rho} = r = \frac{\hat{V}_{xy}}{s_x s_y} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\left(\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{k=1}^{n}(y_k - \overline{y})^2\right)^{1/2}} = \frac{\overline{xy} - \overline{x}\cdot\overline{y}}{\sqrt{\left(\overline{x^2} - \overline{x}^2\right)\left(\overline{y^2} - \overline{y}^2\right)}}$$

r有偏倚。但是当$n \to \infty$时，该偏倚趋于零。

一般而言，概率密度$g(r;\rho,n)$形式复杂；对于高斯变量x, y

$$E[r] = \rho - \frac{\rho(1-\rho^2)}{2n} + \mathcal{O}(n^{-2}); \qquad V[r] = \frac{1}{n}(1-\rho^2)^2 + \mathcal{O}(n^{-2})$$

总结

- ## 估计量

估计量是数据样本的函数$\hat{\theta}(\vec{x})$

评估估计量的好坏：相合性、偏倚性、有效性

- ## 样本均值

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

- ## 样本方差

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

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

- ## 样本协方差

$$\hat{V}_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$