点估计的评价

续本达

清华大学 工程物理系

2023-11-20

复习

点估计 estimator

点估计

用一个数值作为未知参数的估计值 (estimate)

设总体 X 的分布函数的形式已知，θ 是待估参数，$X_1, X_2, \cdots, X_n$ 为总体的一个样本，点估计就是要构造一个适当的统计量 $\hat{\theta}(X_1, X_2, \cdots, X_n)$，用它的观察值 $\hat{\theta}(x_1, x_2, \cdots, x_n)$ 作为待估参数的近似值。

定义 (矩估计)

用样本 k 阶矩作为总体 k 阶矩的估计量，建立含待估参数的方程，从而解出待估参数。

似然函数

1. 设 X 为离散型随机变量，其分布律为

$$P(X=x)=p(x,\theta), x=u_1,u_2,\cdots,\theta\in\Theta$$

样本 $X_1,X_2,\cdots,X_n$ 的联合概率分布为

$$P(X_1=x_1,\cdots,X_n=x_n)=p(x_1,\theta)p(x_2,\theta)\cdots p(x_n,\theta)=L(\vec{x},\theta)$$

2. 设 X 是连续型变量，取 $f(X,\theta)$ 为 X 的密度函数，似然函数定义为

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

最大似然法：maximum likelihood estimation (MLE)

$$\hat{\theta}(\vec{x})=\arg\max L(\vec{x};\theta)$$

称为 最大似然估计值 。称统计量 $\hat{\theta}(X_1,X_2,\cdots,X_n)$ 为参数 θ 的 最大似然估计量 。

评价标准

引子

对于同一个未知参数，不同的方法得到的估计量可能不同，于是提出问题：

1. 应该选用哪一种估计量？
2. 用何标准来评价一个估计量的好坏？

预告：评判标准与数字特征紧密相关

估计量判别	数字特征
无偏性	数学期望
有效性	方差
相合性	依概率收敛

引子

对于同一个未知参数，不同的方法得到的估计量可能不同，于是提出问题：

1. 应该选用哪一种估计量?
2. 用何标准来评价一个估计量的好坏?

预告：评判标准与数字特征紧密相关

估计量判别	数字特征
无偏性	数学期望
有效性	方差
相合性	依概率收敛

相合性

定义

相合性 consistency 又称 一致性

定义

设 $\hat{\theta}(X_1, X_2, \cdots, X_n)$ 是总体参数 θ 的估计量．若

$$\forall\theta \in \Theta(\hat{\theta} \xrightarrow{P} \theta)$$

则称 $\hat{\theta}$ 是 θ 的 相合估计量 .

估计量 $\hat{\theta}$ 实际上依赖于样本容量 n，可以将 $\hat{\theta}_n$ 看作随机变量序列，相合性要求 $\hat{\theta}_n$ 依概率收敛于真值 θ。相合性是好的估计量的基本要求。若不具备相合性，则无论样本容量 n 取多大，都无法将估计量估计得足够精确。

定义

相合性 consistency 又称 一致性

定义

设 $\hat{\theta}(X_1, X_2, \cdots, X_n)$ 是总体参数 θ 的估计量. 若

$$\forall \theta \in \Theta (\hat{\theta} \xrightarrow{P} \theta)$$

则称 $\hat{\theta}$ 是 θ 的 相合估计量 .

估计量 $\hat{\theta}$ 实际上依赖于样本容量 n，可以将 $\hat{\theta}_n$ 看作随机变量序列，相合性要求 $\hat{\theta}_n$ 依概率收敛于真值 θ。相合性是好的估计量的基本要求。若不具备相合性，则无论样本容量 n 取多大，都无法将估计量估计得足够精确。

定义

相合性 consistency 又称 一致性

定义

设 $\hat{\theta}(X_1, X_2, \cdots, X_n)$ 是总体参数 θ 的估计量. 若

$$\forall \theta \in \Theta(\hat{\theta} \xrightarrow{P} \theta)$$

则称 $\hat{\theta}$ 是 θ 的 相合估计量 .

估计量 $\hat{\theta}$ 实际上依赖于样本容量 n，可以将 $\hat{\theta}_n$ 看作随机变量序列，相合性要求 $\hat{\theta}_n$ 依概率收敛于真值 θ。相合性是好的估计量的基本要求。若不具备相合性，则无论样本容量 n 取多大，都无法将估计量估计得足够精确。

相合性判定

定理 (期望方差)

设 $\hat{\theta}_n(X_1, \cdots, X_n)$ 是 θ 的一个估计量。若

$$\lim_{n\to\infty} \mathrm{E}(\hat{\theta}_n) = \theta, \lim_{n\to\infty} \mathrm{Var}(\hat{\theta}_n) = 0$$

则 $\hat{\theta}_n$ 是 θ 的相合估计量。

定义 (均方误差 mean-square error, MSE)

$$\mathrm{MSE}(\hat{\theta}_n) = \mathrm{E}(\hat{\theta}_n - \theta)^2 = \mathrm{Var}(\hat{\theta}_n) + [\mathrm{E}(\hat{\theta}_n) - \theta]^2$$

定理条件可以写为 $\lim_{n\to\infty} \mathrm{MSE}(\hat{\theta}_n) = 0$

相合性判定

定理 (期望方差)

设 $\hat{\theta}_n(X_1, \cdots, X_n)$ 是 θ 的一个估计量。若

$$\lim_{n\to\infty} \mathrm{E}(\hat{\theta}_n) = \theta, \lim_{n\to\infty} \mathrm{Var}(\hat{\theta}_n) = 0$$

则 $\hat{\theta}_n$ 是 θ 的相合估计量。

定义 (均方誤差 mean-square error, MSE)

$$\mathrm{MSE}(\hat{\theta}_n) = \mathrm{E}(\hat{\theta}_n - \theta)^2 = \mathrm{Var}(\hat{\theta}_n) + [\mathrm{E}(\hat{\theta}_n) - \theta]^2$$

定理条件可以写为 $\lim_{n\to\infty} \mathrm{MSE}(\hat{\theta}_n) = 0$

相合估计量函数

定理 (函数的相合性)

若 $\hat{\theta}_{n1},\cdots,\hat{\theta}_{nk}$ 分别是 $\theta_1,\cdots,\theta_k$ 的相合估计量，$\eta = g(\theta_1,\cdots,\theta_k)$ 是 $\theta_1,\cdots,\theta_k$ 的连续函数，则 $\hat{\eta} = g(\hat{\theta}_{n1},\cdots,\hat{\theta}_{nk})$ 是 η 的相合估计量。

例：均匀分布

例

设 $X_1,\cdots,X_n$ 是抽样自总体 $X\sim U(0,\theta)$ 的样本。证明 θ 的最大似然估计量是相合估计量。

θ 的最大似然估计量为 $\hat{\theta}_n=\max(X_1,\cdots,X_n)$。需要证明 $\hat{\theta}_n$

$$\lim_{n\to\infty}\mathrm{E}(\hat{\theta}_n)=\theta,\ \lim_{n\to\infty}\mathrm{Var}(\hat{\theta}_n)=0$$

需要知道 $\hat{\theta}_n$ 服从的分布函数。由独立同分布的随机变量的最大值的分布知，$\hat{\theta}_n$ 服从分布 $F_Y(y)=[F(y)]^n$，其中

$$F(y)=\begin{cases}0, & y<0\\ \frac{y}{\theta}, & 0<y<\theta\\ 1, & y>\theta\end{cases}$$

例：均匀分布

例

设 $X_1, \cdots, X_n$ 是抽样自总体 $X \sim U(0, \theta)$ 的样本。证明 θ 的最大似然估计量是相合估计量。

θ 的最大似然估计量为 $\hat{\theta}_n = \max(X_1, \cdots, X_n)$。需要证明 $\hat{\theta}_n$

$$\lim_{n\to\infty} \mathrm{E}(\hat{\theta}_n) = \theta, \ \lim_{n\to\infty} \mathrm{Var}(\hat{\theta}_n) = 0$$

需要知道 $\hat{\theta}_n$ 服从的分布函数。由独立同分布的随机变量的最大值的分布知，$\hat{\theta}_n$ 服从分布 $F_Y(y) = [F(y)]^n$，其中

$$F(y) = \begin{cases} 0, & y < 0 \\ \frac{y}{\theta}, & 0 < y < \theta \\ 1, & y > \theta \end{cases}$$

例：均匀分布

例

设 $X_1, \cdots, X_n$ 是抽样自总体 $X \sim U(0, \theta)$ 的样本。证明 θ 的最大似然估计量是相合估计量。

θ 的最大似然估计量为 $\hat{\theta}_n = \max(X_1, \cdots, X_n)$。需要证明 $\hat{\theta}_n$

$$\lim_{n\to\infty} \mathrm{E}(\hat{\theta}_n) = \theta, \lim_{n\to\infty} \mathrm{Var}(\hat{\theta}_n) = 0$$

需要知道 $\hat{\theta}_n$ 服从的分布函数。由独立同分布的随机变量的最大值的分布知，$\hat{\theta}_n$ 服从分布 $F_Y(y) = [F(y)]^n$，其中

$$F(y) = \begin{cases} 0, & y < 0 \\ \frac{y}{\theta}, & 0 < y < \theta \\ 1, & y > \theta \end{cases}$$

例：均匀分布

例

设 $X_1, \cdots, X_n$ 是抽样自总体 $X \sim U(0, \theta)$ 的样本。证明 θ 的最大似然估计量是相合估计量。

θ 的最大似然估计量为 $\hat{\theta}_n = \max(X_1, \cdots, X_n)$。需要证明 $\hat{\theta}_n$

$$\lim_{n\to\infty} \mathrm{E}(\hat{\theta}_n) = \theta, \lim_{n\to\infty} \mathrm{Var}(\hat{\theta}_n) = 0$$

需要知道 $\hat{\theta}_n$ 服从的分布函数。由独立同分布的随机变量的最大值的分布知，$\hat{\theta}_n$ 服从分布 $F_Y(y) = [F(y)]^n$，其中

$$F(y) = \begin{cases} 0, & y < 0 \\ \frac{y}{\theta}, & 0 < y < \theta \\ 1, & y > \theta \end{cases}$$

均匀分布：求估计量的期望与方差

$\hat{\theta}_n$ 分布的概率密度函数

$$f_Y(y) = \begin{cases} \frac{ny^{n-1}}{\theta^n}, & 0 < y < \theta \\ 0, & \text{其他} \end{cases}$$

$$\mathrm{E}(\hat{\theta}_n) = \int_0^\theta y \frac{ny^{n-1}}{\theta^n} \mathrm{d}y = \frac{n}{n+1}\theta \xrightarrow{n\to\infty} \theta$$

$$\mathrm{E}(\hat{\theta}_n^2) = \int_0^\theta y^2 \frac{ny^{n-1}}{\theta^n} \mathrm{d}y = \frac{n}{n+2}\theta^2$$

$$\mathrm{Var}(\hat{\theta}_n) = \mathrm{E}(\hat{\theta}_n^2) - \mathrm{E}^2(\hat{\theta}_n) = \frac{n}{(n+1)^2(n+2)}\theta^2 \xrightarrow{n\to\infty} 0$$

所以 $\hat{\theta}_n$ 是相合估计量。

均匀分布：求估计量的期望与方差

$\hat{\theta}_n$ 分布的概率密度函数

$$f_Y(y) = \begin{cases} \frac{ny^{n-1}}{\theta^n}, & 0 < y < \theta \\ 0, & \text{其他} \end{cases}$$

$$\mathrm{E}(\hat{\theta}_n) = \int_0^\theta y \frac{ny^{n-1}}{\theta^n} \mathrm{d}y = \frac{n}{n+1}\theta \xrightarrow{n\to\infty} \theta$$

$$\mathrm{E}(\hat{\theta}_n^2) = \int_0^\theta y^2 \frac{ny^{n-1}}{\theta^n} \mathrm{d}y = \frac{n}{n+2}\theta^2$$

$$\mathrm{Var}(\hat{\theta}_n) = \mathrm{E}(\hat{\theta}_n^2) - \mathrm{E}^2(\hat{\theta}_n) = \frac{n}{(n+1)^2(n+2)}\theta^2 \xrightarrow{n\to\infty} 0$$

所以 $\hat{\theta}_n$ 是相合估计量。

均匀分布：求估计量的期望与方差

$\hat{\theta}_n$ 分布的概率密度函数

$$f_Y(y) = \begin{cases} \frac{ny^{n-1}}{\theta^n}, & 0 < y < \theta \\ 0, & \text{其他} \end{cases}$$

$$\mathrm{E}(\hat{\theta}_n) = \int_0^\theta y \frac{ny^{n-1}}{\theta^n} \mathrm{d}y = \frac{n}{n+1}\theta \xrightarrow{n\to\infty} \theta$$

$$\mathrm{E}(\hat{\theta}_n^2) = \int_0^\theta y^2 \frac{ny^{n-1}}{\theta^n} \mathrm{d}y = \frac{n}{n+2}\theta^2$$

$$\mathrm{Var}(\hat{\theta}_n) = \mathrm{E}(\hat{\theta}_n^2) - \mathrm{E}^2(\hat{\theta}_n) = \frac{n}{(n+1)^2(n+2)}\theta^2 \xrightarrow{n\to\infty} 0$$

所以 $\hat{\theta}_n$ 是相合估计量。

均匀分布：求估计量的期望与方差

$\hat{\theta}_n$ 分布的概率密度函数

$$f_Y(y) = \begin{cases} \frac{ny^{n-1}}{\theta^n}, & 0 < y < \theta \\ 0, & \text{其他} \end{cases}$$

$$\mathrm{E}(\hat{\theta}_n) = \int_0^\theta y \frac{ny^{n-1}}{\theta^n} \mathrm{d}y = \frac{n}{n+1}\theta \xrightarrow{n\to\infty} \theta$$

$$\mathrm{E}(\hat{\theta}_n^2) = \int_0^\theta y^2 \frac{ny^{n-1}}{\theta^n} \mathrm{d}y = \frac{n}{n+2}\theta^2$$

$$\mathrm{Var}(\hat{\theta}_n) = \mathrm{E}(\hat{\theta}_n^2) - \mathrm{E}^2(\hat{\theta}_n) = \frac{n}{(n+1)^2(n+2)}\theta^2 \xrightarrow{n\to\infty} 0$$

所以 $\hat{\theta}_n$ 是相合估计量。

无偏性

定义 unbiased

定义

若 $\mathrm{E}(\hat{\theta}) = \theta$ 则称 $\hat{\theta}$ 是 θ 的无偏估计量。

我们不可能要求每一次由样本得到的估计量与真值都相等，但可以要求这些估计量的期望与真值相等.

- 样本均值 $\overline{X}$ 是总体期望 $\mathrm{E}(X)$ 的 无偏估计量 。
- 样本二阶原点矩 $A_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ 是总体二阶原点矩 $\mu_2 = \mathrm{E}(X^2)$ 的 无偏估计量 。
- $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ 是 $\mathrm{Var}(X)$ 的 无偏估计
- $S_n^2 = \frac{n-1}{n} S^2$ 是 $\mathrm{Var}(X)$ 的 渐进无偏估计

定义 unbiased

定义

若 $\mathrm{E}(\hat{\theta}) = \theta$ 则称 $\hat{\theta}$ 是 θ 的无偏估计量。

我们不可能要求每一次由样本得到的估计量与真值都相等，但可以要求这些估计量的期望与真值相等.

- 样本均值 $\overline{X}$ 是总体期望 $\mathrm{E}(X)$ 的 无偏估计量 。
- 样本二阶原点矩 $A_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ 是总体二阶原点矩 $\mu_2 = \mathrm{E}(X^2)$ 的 无偏估计量 。
- $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ 是 $\mathrm{Var}(X)$ 的 无偏估计 .
- $S_n^2 = \frac{n-1}{n} S^2$ 是 $\mathrm{Var}(X)$ 的 渐进无偏估计 .

定义 unbiased

定义

若 $\mathrm{E}(\hat{\theta}) = \theta$ 则称 $\hat{\theta}$ 是 θ 的无偏估计量。

我们不可能要求每一次由样本得到的估计量与真值都相等，但可以要求这些估计量的期望与真值相等.

- 样本均值 $\overline{X}$ 是总体期望 $\mathrm{E}(X)$ 的 无偏估计量 。
- 样本二阶原点矩 $A_2 = \dfrac{1}{n}\sum_{i=1}^{n} X_i^2$ 是总体二阶原点矩 $\mu_2 = \mathrm{E}(X^2)$ 的 无偏估计量 。
- $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ 是 $\mathrm{Var}(X)$ 的 无偏估计 .
- $S_n^2 = \dfrac{n-1}{n} S^2$ 是 $\mathrm{Var}(X)$ 的 渐进无偏估计 .

定义 unbiased

定义

若 $\mathrm{E}(\hat{\theta}) = \theta$ 则称 $\hat{\theta}$ 是 θ 的无偏估计量。

我们不可能要求每一次由样本得到的估计量与真值都相等，但可以要求这些估计量的期望与真值相等.

- 样本均值 $\overline{X}$ 是总体期望 $\mathrm{E}(X)$ 的 无偏估计量 。
- 样本二阶原点矩 $A_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ 是总体二阶原点矩 $\mu_2 = \mathrm{E}(X^2)$ 的 无偏估计量 。
- $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ 是 $\mathrm{Var}(X)$ 的 无偏估计 .
- $S_n^2 = \frac{n-1}{n} S^2$ 是 $\mathrm{Var}(X)$ 的 渐进无偏估计 .

定义 unbiased

定义

若 $\mathrm{E}(\hat{\theta}) = \theta$ 则称 $\hat{\theta}$ 是 θ 的无偏估计量。

我们不可能要求每一次由样本得到的估计量与真值都相等，但可以要求这些估计量的期望与真值相等.

- 样本均值 $\overline{X}$ 是总体期望 $\mathrm{E}(X)$ 的 无偏估计量 。
- 样本二阶原点矩 $A_2 = \dfrac{1}{n}\sum_{i=1}^{n} X_i^2$ 是总体二阶原点矩 $\mu_2 = \mathrm{E}(X^2)$ 的 无偏估计量 。
- $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ 是 $\mathrm{Var}(X)$ 的 无偏估计 .
- $S_n^2 = \dfrac{n-1}{n}S^2$ 是 $\mathrm{Var}(X)$ 的 渐进无偏估计 .

释义：相合性与无偏性

回顾相合性判定定理

定理 (期望方差)

设 $\hat{\theta}_n(X_1, \cdots, X_n)$ 是 θ 的一个估计量。若

$$\lim_{n\to\infty} \mathrm{E}(\hat{\theta}_n) = \theta, \lim_{n\to\infty} \mathrm{Var}(\hat{\theta}_n) = 0$$

则 $\hat{\theta}_n$ 是 θ 的相合估计量。

前一个条件正是 $\hat{\theta}_n$ 的渐进无偏性。

相合性与（渐进）无偏性的区别

相合性不能推出渐进无偏性，只有渐进无偏性也不能推出相合性。

释义：相合性与无偏性

回顾相合性判定定理

定理 (期望方差)

设 $\hat{\theta}_n(X_1,\cdots,X_n)$ 是 θ 的一个估计量。若

$$\lim_{n\to\infty} \mathrm{E}(\hat{\theta}_n) = \theta, \lim_{n\to\infty} \mathrm{Var}(\hat{\theta}_n) = 0$$

则 $\hat{\theta}_n$ 是 θ 的相合估计量。

前一个条件正是 $\hat{\theta}_n$ 的渐进无偏性。

相合性与（渐进）无偏性的区别

相合性不能推出渐进无偏性，只有渐进无偏性也不能推出相合性。

例：二项分布

例

设 $(X_1, X_2, \cdots, X_m)$ 是总体 $X \sim B(n,p)$ 的一个样本，$n > 1$，求 p^2 的无偏估计量.

$$\mathrm{E}(\overline{X}) = \mathrm{E}(X) = np$$

$$\mathrm{E}\left(\frac{1}{m}\sum_{i=1}^{m} X_i^2\right) = E(X^2) = (np)^2 + np(1-p) = (n^2-n)p^2 + \mathrm{E}(\overline{X})$$

$$\Longrightarrow \quad p^2 = \frac{1}{n^2-n}\left[\mathrm{E}\left(\frac{1}{m}\sum_{i=1}^{m} X_i^2\right) - \overline{X}\right] = \mathrm{E}\left[\frac{1}{(n^2-n)m}\sum_{i=1}^{m} X_i(X_i-1)\right]$$

$$\Longrightarrow \quad \widehat{p^2} = \frac{1}{(n^2-n)m}\sum_{i=1}^{m} X_i(X_i-1)$$

是无偏估计量。

例：二项分布

例

设 $(X_1, X_2, \cdots, X_m)$ 是总体 $X \sim B(n,p)$ 的一个样本，$n > 1$，求 p^2 的无偏估计量.

$$\mathrm{E}(\overline{X}) = \mathrm{E}(X) = np$$

$$\mathrm{E}\left(\frac{1}{m}\sum_{i=1}^{m} X_i^2\right) = E(X^2) = (np)^2 + np(1-p) = (n^2-n)p^2 + \mathrm{E}(\overline{X})$$

$$\Longrightarrow\ p^2 = \frac{1}{n^2-n}\left[\mathrm{E}\left(\frac{1}{m}\sum_{i=1}^{m} X_i^2\right) - \overline{X}\right] = \mathrm{E}\left[\frac{1}{(n^2-n)m}\sum_{i=1}^{m} X_i(X_i-1)\right]$$

$$\Longrightarrow\ \widehat{p^2} = \frac{1}{(n^2-n)m}\sum_{i=1}^{m} X_i(X_i-1)$$

是无偏估计量。

例：二项分布

例

设 $(X_1, X_2, \cdots, X_m)$ 是总体 $X \sim B(n, p)$ 的一个样本，$n > 1$，求 p^2 的无偏估计量.

$$\mathrm{E}(\overline{X}) = \mathrm{E}(X) = np$$

$$\mathrm{E}\left(\frac{1}{m}\sum_{i=1}^{m} X_i^2\right) = E(X^2) = (np)^2 + np(1-p) = (n^2-n)p^2 + \mathrm{E}(\overline{X})$$

$$\Longrightarrow\ p^2 = \frac{1}{n^2-n}\left[\mathrm{E}\left(\frac{1}{m}\sum_{i=1}^{m} X_i^2\right) - \overline{X}\right] = \mathrm{E}\left[\frac{1}{(n^2-n)m}\sum_{i=1}^{m} X_i(X_i-1)\right]$$

$$\Longrightarrow\ \hat{p^2} = \frac{1}{(n^2-n)m}\sum_{i=1}^{m} X_i(X_i-1)$$

是无偏估计量。

例：二项分布

例

设 $(X_1, X_2, \cdots, X_m)$ 是总体 $X \sim B(n,p)$ 的一个样本，$n > 1$，求 p^2 的无偏估计量.

$$\mathrm{E}(\overline{X}) = \mathrm{E}(X) = np$$

$$\mathrm{E}\left(\frac{1}{m}\sum_{i=1}^{m} X_i^2\right) = E(X^2) = (np)^2 + np(1-p) = (n^2-n)p^2 + \mathrm{E}(\overline{X})$$

$$\Longrightarrow\ p^2 = \frac{1}{n^2-n}\left[\mathrm{E}\left(\frac{1}{m}\sum_{i=1}^{m} X_i^2\right) - \overline{X}\right] = \mathrm{E}\left[\frac{1}{(n^2-n)m}\sum_{i=1}^{m} X_i(X_i-1)\right]$$

$$\Longrightarrow\ \hat{p^2} = \frac{1}{(n^2-n)m}\sum_{i=1}^{m} X_i(X_i-1)$$

是无偏估计量。

有效性

定义

定义

设 $\hat{\theta}_1(X_1, X_2, \cdots, X_n)$ 和 $\hat{\theta}_2(X_1, X_2, \cdots, X_n)$ 都是总体参数 θ 的无偏估计量，且 $\mathrm{Var}(\hat{\theta}_1) < \mathrm{Var}(\hat{\theta}_2)$ 则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 更 有效 .

一致最小方差无偏估计 UMVUE

如果一个估计量比任何其它估计量都有效，则称之为 一致最小方差无偏估计 uniformly minimum variance unbiased estimator UMVUE

定义

定义

设 $\hat{\theta}_1(X_1, X_2, \cdots, X_n)$ 和 $\hat{\theta}_2(X_1, X_2, \cdots, X_n)$ 都是总体参数 θ 的无偏估计量, 且 $\mathrm{Var}(\hat{\theta}_1) < \mathrm{Var}(\hat{\theta}_2)$ 则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 更 有效 .

一致最小方差无偏估计 UMVUE

如果一个估计量比任何其它估计量都有效，则称之为 一致最小方差无偏估计 uniformly minimum variance unbiased estimator UMVUE

估计量的方差极限

定理 (Cramer-Rao 下限)

$L(X;\theta)$ 是带参数 θ 的总体 X 的似然函数。任何 θ 的无偏估计量 $\hat{\theta}_n$ 满足

$$\mathrm{Var}(\hat{\theta}_n) \geq \frac{1}{n\,\mathrm{E}\left(\frac{\partial \log L(X;\theta)}{\partial \theta}\right)^2}$$

达到 Cramer-Rao 理论极限的估计量称为 有效估计量 。

例：最大似然估计量

以下重要结论是 最大似然估计量 应用的理论支撑。证明较难，不做细致讨论

1. 如果参数存在有效无偏估计量，那么它一定是最大似然估计量
2. 一般情况下，最大似然估计量是一致的
3. 最大似然估计量渐进服从正态分布

例：最大似然估计量

以下重要结论是 最大似然估计量 应用的理论支撑。证明较难，不做细致讨论

1. 如果参数存在有效无偏估计量，那么它一定是最大似然估计量
2. 一般情况下，最大似然估计量是一致的
3. 最大似然估计量渐进服从正态分布

例：最大似然估计量

以下重要结论是 最大似然估计量 应用的理论支撑。证明较难，不做细致讨论

1. 如果参数存在有效无偏估计量，那么它一定是最大似然估计量
2. 一般情况下，最大似然估计量是一致的
3. 最大似然估计量渐进服从正态分布

例：信息积累

例

设 $X_1, \cdots, X_n$ 是取自某总体的样本，记总体均值和方差分别为 μ 和 σ^2。μ 的两个估计量分别定义为 $\hat{\mu}_1 = X_1$，$\hat{\mu}_2 = \overline{X}$。哪个估计量更有效?

$$E(\hat{\mu}_1) = E(X_1) = \mu, E(\hat{\mu}_2) = E(\overline{X}) = \mu$$

两者都无偏。

$$\mathrm{Var}(\hat{\mu}_1) = \sigma^2, \mathrm{Var}(\hat{\mu}_2) = \frac{\sigma^2}{n}$$

当 $n > 1$ 时，$\mathrm{Var}(\hat{\mu}1) < \mathrm{Var}(\hat{\mu}_2)$，$\hat{\mu}_2$ 更有效。

启示

1. 用全部数据的平均估计总体均值要比只使用部分数据更有效。
2. 数据积累，估计量的方差越来越小，参数越来越精确，信息量越来越大。

例：信息积累

例

设 $X_1, \cdots, X_n$ 是取自某总体的样本，记总体均值和方差分别为 μ 和 σ^2。μ 的两个估计量分别定义为 $\hat{\mu}_1 = X_1$，$\hat{\mu}_2 = \overline{X}$。哪个估计量更有效?

$$E(\hat{\mu}_1) = E(X_1) = \mu, E(\hat{\mu}_2) = E(\overline{X}) = \mu$$

两者都无偏。

$$\mathrm{Var}(\hat{\mu}_1) = \sigma^2, \mathrm{Var}(\hat{\mu}_2) = \frac{\sigma^2}{n}$$

当 $n > 1$ 时，$\mathrm{Var}(\hat{\mu}1) < \mathrm{Var}(\hat{\mu}_2)$，$\hat{\mu}_2$ 更有效。

启示

1. 用全部数据的平均估计总体均值要比只使用部分数据更有效。
2. 数据积累，估计量的方差越来越小，参数越来越精确， 信息量 越来越大。

例：最小二乘估计

背景

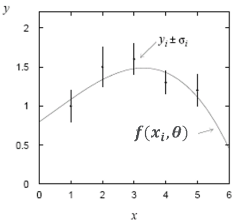
- 最小二乘估计法（Least Square Method, LSE）是自古以来最广泛使用的参数估计法之一，源于天文学和测地学上的应用。
- 当总体分布的函数形式并不严格知道、无法进行最大似然估计时，运用最小二乘法往往十分方便。
- 由 Legendre 和 Gauss 在同一时代发现。

引例

- 假设某个随机变量 Y 与 X 和未知参数 θ 有关：

$$Y = f(X;\theta)$$

- 为了估计参数 θ，在 X 的不同取值 $x_1, \cdots, x_N$ 测量 Y，得到对应的测量值 $y_1, \cdots, y_N$。然后用函数 $y = f(x, \theta)$ “拟合”数据。
- 直观来说，如果得到的“拟合” 曲线与每个数据点距离越小，则该曲线所对应的参数越好。



定义

- “数据拟合”常用实现方法是最小二乘法。

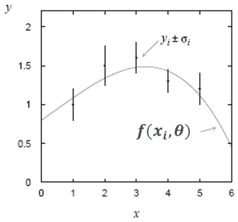
$$\chi^2(\theta) \equiv \sum_{i=1}^{N} [y_i - f(x_i;\theta)]^2$$

取最小值的 $\hat{\theta}$ 为估计量。

- 若在不同 x_i 处测量得到的 y_i 的精度不同，假设 y_i 的方差为 σ_i^2，则 $\chi^2(\theta)$ 定义为

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

y_i 经常近似服从正态分布，则 $\chi^2(\theta)$ 近似服从卡方分布。



最小二乘法与最大似然法的等价性

一定条件下，最小二乘法与最大似然法等价。
假设每次测量独立，观测值 $Y_i \sim N(\eta_i, \sigma_i^2)$，其中真值 η_i 未知。
N 次观测 $\vec{Y} = (Y_1, \cdots, Y_N)$ 的似然函数为

$$L = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(Y_i-\eta_i)^2}{2\sigma_i^2}} \propto e^{-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{Y_i-\eta_i}{\sigma_i}\right)^2}$$

要使 L 取最大，则要 $\sum_{i=1}^{N}(\frac{Y_i-\eta_i}{\sigma_i})^2$ 最小

最小二乘估计的普适性

- 指数族分布 包含正态分布、泊松分布、二项分布、伽马分布
- 在指数族分布的参数估计中， 迭代重加权最小二乘法 等价于最大似然法。

实验物理学家的三个境界

1. 不管随机变量的分布 → 无脑使用最小二乘法
2. 重视随机变量的分布 → 无脑使用最大似然法
3. 理解参数估计理论，上过《概率统计分析及量测技术》 → 批判地选择最有效的方法

最小二乘估计的普适性

- 指数族分布 包含正态分布、泊松分布、二项分布、伽马分布
- 在指数族分布的参数估计中， 迭代重加权最小二乘法 等价于最大似然法。

实验物理学家的三个境界

1. 不管随机变量的分布 → 无脑使用最小二乘法
2. 重视随机变量的分布 → 无脑使用最大似然法
3. 理解参数估计理论，上过《概率统计分析及量测技术》 → 批判地选择最有效的方法

例：截尾分布

量程可能小于随机变量的取值

“卡一个窗口拟合”，需要注意

1. 窗口也是参数。
2. 卡窗口之后，应重新归一。