

法律声明

□ 本课件包括：演示文稿，示例，代码，题库，视频和声音等，小象学院拥有完全知识产权的权利；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或机构不得盗版、复制、仿造其中的创意，我们将保留一切通过法律手段追究违反者的权利。

□ 课程详情请咨询

■ 微信公众号：小象

■ 新浪微博：ChinaHadoop



数理统计

目录

第一章	统计量与抽样分布
第二章	点估计
第三章	区间估计
第四章	假设检验
第五章	分布的检验
第六章	Bayes统计与统计判决理论
第七章	非参数统计
第八章	抽样调查

目 录

2.1 矩估计

2.2 极大似然估计

2.3 无偏估计与一致最小方差无偏估计

2.4 完备统计量

2.5 CR不等式及有效估计

2.6 相合估计

参数估计

参数估计有两种形式：点估计与区间估计。这两种方法互为补充。

本章，我们重点讨论点估计(矩估计和极大似然估计)。并且提出一些列评价估计量优良性的准则：无偏性、有效性、相合性、渐近正态性等。

替换原理与矩法估计

1900年Pearson提出了一个替换原理，后来人们称此法为矩法。

替换原理常指如下两句话：

- (1)用样本矩去替换总体矩，这里的矩可以是原点矩也可以是中心矩。
- (2)用样本矩的函数去替换响应的总体矩的函数。

根据这个替换原理，在总体分布形式未知场合也可以对各种参数作出估计，譬如：

- (1)用样本均值估计总体均值；
- (2)用样本方差估计总体方差；
- (3)用事件A出现的频率估计事件A发生的概率；
- (4)用样本的p分位数估计总体的p分位数。

矩法估计的统计思想(替换原理)十分简单明确，众人都能接受，使用场合很广。它的实质是用经验分布函数去替换总体分布，其理论基础是格列文科定理。

矩估计

定义 2.2.1 对于样本 $X_1, \dots, X_n$ 及任意一正整数 k , 我们称

$$a_k = \frac{1}{n} \sum_{i=1}^n X_i^k, \quad m_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k, \quad (2.2.1)$$

为样本 k 阶原点矩和 k 阶中心矩, 其中 $\bar{X} = a_1$ 为样本均值.

我们在概率论已经讲过一个随机变量的 k 阶原点矩和中心矩, 为与样本矩做一比较, 我们称总体 X 的 k 阶原点矩和中心矩分别为

$$\mu_k = EX^k, \quad \nu_k = E(X - \mu_1)^k. \quad (2.2.2)$$

显然, 样本矩不依赖于总体中的参数, 它们是统计量; 而总体矩则与分布中的未知参数有关. 另外, 由大数定律和中心极限定理可知, 样本矩是总体矩的一个很好的估计. 于是, 我们可以利用上述想法来估计总体的未知参数, 这就是矩估计的基本思想

矩估计

矩法 (method of moments) 是由英国统计学家 K · Pearson 在 20 世纪初提出来的, 它的中心思想就是用样本矩去估计总体矩.

设总体 X 的分布中的未知参数为 $\theta = (\theta_1, \theta_2, \dots, \theta_m)^T$, 假定总体 X 的 k 阶原点矩

$$E(X^k) = \alpha_k(\theta_1, \theta_2, \dots, \theta_m), \quad k = 1, 2, \dots, m$$

存在, 我们令总体的 k 阶原点矩等于它样本的 k 阶原点矩

$$A_k = \frac{1}{n} \sum_{i=1}^n X_i^k, \quad k = 1, 2, \dots, m,$$

即

$$\alpha_k(\theta_1, \theta_2, \dots, \theta_m) = E(X^k) = \frac{1}{n} \sum_{i=1}^n X_i^k = A_k, \quad k = 1, 2, \dots, m. \quad (4.1)$$

矩估计

由方程 (4.1) 可以得到关于未知量 θ 的解

$$\hat{\theta}_i = \hat{\theta}_i(X_1, X_2, \dots, X_n), \quad i = 1, 2, \dots, m. \quad (4.2)$$

取 $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m)^T$ 作为 $\theta = (\theta_1, \theta_2, \dots, \theta_m)^T$ 的估计, 则称 $\hat{\theta}$ 为 θ 的矩估计 (estimation by moments), 用矩估计参数的方法称为矩法.

矩估计

例 4.1 设总体 X 的均值为 μ , 方差为 σ^2 , $X_1, X_2, \dots, X_n$ 是来自总体 X 的一个样本, 试用矩方法估计均值 μ , 和方差 σ^2 .

矩估计

解: 计算总体 X 的一阶、二阶原点矩

$$\alpha_1 = E(X) = \mu,$$

$$\alpha_2 = E(X^2) = \text{Var}(X) + [E(X)]^2 = \sigma^2 + \mu^2.$$

和样本的一阶、二阶原点矩

$$A_1 = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}, \quad A_2 = \frac{1}{n} \sum_{i=1}^n X_i^2.$$

由式 (4.1) 得到方程组

$$\begin{cases} \mu = \bar{X}, \\ \sigma^2 + \mu^2 = \frac{1}{n} \sum_{i=1}^n X_i^2. \end{cases}$$

解上述方程组得到均值 μ 和方差 σ^2 的矩估计

$$\hat{\mu} = \bar{X}, \tag{4.3}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2. \tag{4.4}$$

矩估计

需要特别注意的是：方差的矩估计并不等于样本方差 S^2 ，而是有如下关系式

$$\hat{\sigma}^2 = \frac{n-1}{n} S^2. \quad (4.5)$$

对于正态分布 $N(\mu, \sigma^2)$ ，因为 μ 和 σ^2 分别为总体的均值和方差，由式 (4.3) 和式 (4.4) 得到参数 μ 和 σ^2 的矩估计

$$\hat{\mu} = \bar{X}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

从上述过程，可以看到，利用矩法估计均值和方差，就等价于用样本的一阶原点矩估计均值，用样本的二阶中心矩估计方差。

矩估计

例 4.4 设总体 X 是区间 $[a, b]$ 上的均匀分布, 其中 a, b 是未知参数, $X_1, X_2, \dots, X_n$ 是总体 X 的一个样本, 试用矩估法估计参数 a 和 b .

解: 由例 4.1 的计算过程(式 (4.3)-(4.4))可知, 用一、二阶原点矩作估计, 本质上相当用一阶原点估计均值, 二阶中心矩估计方差, 即

$$E(X) = A_1 = \frac{1}{n} \sum_{i=1}^n X_i, \quad \text{Var}(X) = M_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

均匀分布的均值是 $(b-a)/2$, 方差是 $(b-a)^2/12$, 所以令

$$\frac{b+a}{2} = \bar{X}, \quad \frac{(b-a)^2}{12} = M_2,$$

解上述方程组得到 a 和 b 的估计分别为

$$\hat{a} = \bar{X} - \sqrt{3M_2}, \quad \hat{b} = \bar{X} + \sqrt{3M_2}. \quad (4.6)$$

相合性

定义 4.5 设 $\hat{\theta}(X_1, X_2, \dots, X_n)$ 为未知参数 θ 的估计量, 若对于任意 $\theta \in \Theta$, 当 $n \rightarrow \infty$ 时, $\hat{\theta}(X_1, X_2, \dots, X_n)$ 依概率收敛于 θ , 即对任意 $\varepsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P\{|\hat{\theta} - \theta| < \varepsilon\} = 1,$$

则称 $\hat{\theta}$ 为 θ 的相合估计 (*consistent estimate*) 量或一致估计量, 并记为 $\hat{\theta} \xrightarrow{P} \theta (n \rightarrow \infty)$.

若当 $n \rightarrow \infty$ 时, $\hat{\theta}$ 均方收敛于 θ , 即

$$\lim_{n \rightarrow \infty} E(\hat{\theta} - \theta)^2 = 0,$$

则称 $\hat{\theta}$ 为 θ 的均方相合估计量 (或一致估计量), 并记为 $\hat{\theta} \xrightarrow{L^2} \theta (n \rightarrow \infty)$.

辛钦大数定律

辛钦大数定律: 设 $x_1, x_2, \dots, x_n$ 是一列独立同分布的随机变量序列, 若其数学期望 μ 有限, 则对任意给定的 $\epsilon > 0$, 有

$$P\left(\left|\frac{1}{n} \sum_{i=1}^n x_i - \mu\right| > \epsilon\right) \rightarrow 0 (n \rightarrow \infty)$$

相合性

定理: 设 $\hat{\theta}_{n1}, \hat{\theta}_{n2}, \dots, \hat{\theta}_{nk}$ 分别是 $\theta_1, \theta_2, \dots, \theta_k$ 的相合估计, 若 $g(\theta_1, \theta_2, \dots, \theta_k)$ 是 k 元连续函数, 则 $\hat{g}_n = g(\hat{\theta}_{n1}, \hat{\theta}_{n2}, \dots, \hat{\theta}_{nk})$ 是 $g(\theta_1, \theta_2, \dots, \theta_k)$ 的相合估计.

相合性

定理: 设 $\hat{\theta}_n = \hat{\theta}_n(x_1, \dots, x_n)$ 是 θ 的一个估计量, 若

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

则 $\hat{\theta}_n$ 是 θ 的相合估计.

目 录

2.1 矩估计

2.2 极大似然估计

2.3 无偏估计与一致最小方差无偏估计

2.4 完备统计量

2.5 CR不等式及有效估计

2.6 相合估计

极大似然原理

例 2.3.1 假设在一个罐中放着许多白球和黑球，且二者数目之比为1:3，但不知哪种颜色球多。现考虑通过返回抽样方法估计罐中黑球数目的比例 p 。

解 假设从罐中返回抽样 n 次，则由概率论知识知道，其中包括黑球个数 X 的概率分布是二项分布 $B(n, p)$ ，即

$$P(x, p) = P\{X = x|p\} = \binom{n}{x} p^x (1 - p)^{n-x}. \quad (2.3.1)$$

由于我们知道 p 等于1/4或3/4，故我们的目的在于通过 X 的大小来估计 p 。为此，我们先看一个 $n = 3$ 的简单情况。此时，我们针对不同的 x 值，计算(2.3.1)式的概率如下：

表2.1 极大似然思想

x	0	1	2	3
$P(x, \frac{3}{4})$	$\frac{1}{64}$	$\frac{9}{64}$	$\frac{27}{64}$	$\frac{27}{64}$
$P(x, \frac{1}{4})$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

极大似然原理

从上表可以看出, 当 $x = 0$ 时, 因为 $\frac{27}{64} \gg \frac{1}{64}$, 故认为 $\hat{p} = \frac{1}{4}$ 更有可能性发生, 因而, $\hat{p} = \frac{1}{4}$ 比取 $\hat{p} = \frac{3}{4}$ 更合理. 于是, 如把 $\hat{p}$ 看成是 x 的函数 $\hat{p}(x)$, 则我们得到如下的比较合理的估计:

$$\hat{p}(x) = \begin{cases} \frac{1}{4}, & \text{当 } x = 0, 1, \\ \frac{3}{4}, & \text{当 } x = 2, 3. \end{cases}$$

我们在选取上述估计时遵循了如下原理: 认为概率最大的事件最有可能发生, 即我们选取的估计 $\hat{p}(x)$ 满足如下不等式:

$$P\{X = x|\hat{p}(x)\} \geq P\{X = x|p'\}, \quad \forall p', \forall x.$$

这就是我们要讲的极大似然原理的基本思想. □

极大似然原理

极大似然法是 Fisher(费希尔) 在 1912 年提出的一种应用非常广泛的参数估计方法, 其思想始于 Gauss 的误差理论, 它具有很多优良的性质. 它充分利用总体分布函数的信息, 克服了矩法的某些不足.

设 Θ 是参数空间, 参数 θ 可取 Θ 的所有值, 在给定样本的观察值 $(x_1, x_2, \dots, x_n)$ 后, 不同的 θ 对应于 $(X_1, X_2, \dots, X_n)$ 落入 $(x_1, x_2, \dots, x_n)$ 的邻域内的概率大小不同, 既然在一次试验中就观察到了 $(X_1, X_2, \dots, X_n)$ 的取值为 $(x_1, x_2, \dots, x_n)$, 因此, 可以认为 θ 是最有可能来源于使 $(X_1, X_2, \dots, X_n)$ 落入 $(x_1, x_2, \dots, x_n)$ 邻域内的概率达到最大者 $\hat{\theta}$, 即

$$\prod_{i=1}^n f(x_i; \hat{\theta}) = \sup_{\theta \in \Theta} \prod_{i=1}^n f(x_i; \theta). \quad (4.8)$$

取 $\hat{\theta}$ 作为 θ 的估计, 这就是极大似然原理.

似然函数

定义 4.1 设总体 X 的概率密度函数或分布律为 $f(x; \theta)$, $\theta \in \Theta$ 是未知参数, $X_1, X_2, \dots, X_n$ 来自总体 X 的样本, 称

$$L(\theta; x) = L(\theta; x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i; \theta)$$

为 θ 的似然函数 (*likelihood function*).

显然, 若样本取值 x 固定时, $L(\theta; x)$ 是 θ 的函数. 若参数 θ 固定, 当 X 为连续型随机变量时, 它就是样本 $(X_1, X_2, \dots, X_n)$ 的联合概率密度函数; 当 X 为离散型随机变量时, 它就是样本 $(X_1, X_2, \dots, X_n)$ 的联合分布律.

极大似然估计

定义 4.2 设总体 X 的概率密度函数或分布律为 $f(x; \theta)$, $\theta \in \Theta$ 是未知参数, $X_1, X_2, \dots, X_n$ 来自总体 X 的样本, $L(\theta; x)$ 为 θ 的似然函数, 若 $\hat{\theta} = \hat{\theta}(X) = \hat{\theta}(X_1, X_2, \dots, X_n)$ 是一个统计量且满足

$$L(\hat{\theta}(X); X) = \sup_{\theta \in \Theta} L(\theta; X),$$

则称 $\hat{\theta}(X)$ 为 θ 的极大似然估计 (*maximum likelihood estimation*), 简记为 *MLE*.
用极大似然估计来估计参数的方法为称极大似然法.

极大似然估计

下面分不同情况介绍极大似然法的求解过程.

(1) 似然函数 $L(\theta; X)$ 为 θ 的连续函数, 且关于 θ 的各分量的偏导数存在.

设 θ 是 m 维变量, 且 $\Theta \subset R^m$ 为开区域, 则由极值的一阶必要条件, 得到

$$\frac{\partial L(\theta; X)}{\partial \theta_i} = 0, \quad i = 1, 2, \dots, m. \quad (4.9)$$

通常称式 (4.9) 为似然方程. 由于独立同分布的样本的似然函数 $L(\theta; X)$ 具有连乘积的形式, 故对 $L(\theta; X)$ 取对数后再求偏导数是方便的, 因此实用上常采用与 (4.9) 等价的形式

$$\frac{\partial \ln L(\theta; X)}{\partial \theta_i} = 0, \quad i = 1, 2, \dots, m. \quad (4.10)$$

称式 (4.10) 为对数似然方程 (loglikelihood equation).

极大似然估计

例 4.6 设总体 X 服从正态分布 $N(\mu, \sigma^2)$, 其中 μ, σ^2 为未知参数, $X_1, X_2, \dots, X_n$ 是来自总体 X 的一个样本, 试用极大似然法估计参数 (μ, σ^2) .

极大似然估计

解：正态分布的似然函数为

$$L(\mu, \sigma^2; x) = \prod_{i=1}^n f(x_i; \mu, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right],$$

相应的对数似然函数为

$$\ln L(\mu, \sigma^2; x) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

令

$$\begin{cases} \frac{\partial \ln L(\mu, \sigma^2; x)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0, \\ \frac{\partial \ln L(\mu, \sigma^2; x)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 = 0, \end{cases}$$

解此似然方程组得到：

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}, \quad \sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2.$$

极大似然估计

进一步验证, 对于对数似然函数 $\ln L(\mu, \sigma^2; x)$ 的二阶 Hesse 矩阵

$$\begin{bmatrix} -\frac{n}{\sigma^2} & -\frac{1}{\sigma^4} \sum_{i=1}^n (x_i - \mu) \\ -\frac{1}{\sigma^4} \sum_{i=1}^n (x_i - \mu) & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (x_i - \mu)^2 \end{bmatrix} = \begin{bmatrix} -\frac{n}{\sigma^2} & 0 \\ 0 & -\frac{n}{2\sigma^4} \end{bmatrix}$$

是负定矩阵, 所以 $\left(\bar{x}, \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)$ 是 $L(\mu, \sigma^2; x)$ 的极大值点. 故 (μ, σ^2) 的极大似然估计是

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

与例 4.1 相比较, 两者的计算结果是相同的.

极大似然估计

(2) 似然函数 $L(\theta; x)$ 关于 θ 有间断点.

当 Θ 为 R^m 中的开区域, 此时求似然方程组解的方法不适用, 要具体问题具体分析.

例 4.8 设总体 X 是区间 $[a, b]$ 上的均匀分布, 其中 a, b 是未知参数, $X_1, X_2, \dots, X_n$ 是总体 X 的一个样本, 试用极大似然法估计参数 a 和 b .

极大似然估计

解: 对于样本 $X_1, X_2, \dots, X_n$, 其似然函数为

$$L(a, b; x) = \begin{cases} \frac{1}{(b-a)^n}, & \text{若 } a \leq x_i \leq b, \quad i = 1, 2, \dots, n, \\ 0, & \text{其它.} \end{cases}$$

很显然, $L(a, b; x)$ 不是 (a, b) 的连续函数, 因此不能用似然方程组 (4.10) 求解, 而必需从极大似然估计的定义出发来求 $L(a, b; x)$ 的最大值. 为了使 $L(a, b; x)$ 达到最大, 则 $b - a$ 应该尽可能的小, 但 b 不能小于 $\max\{x_1, x_2, \dots, x_n\}$; 否则 $L(a, b; x) = 0$. 类似地, a 不能大于 $\min\{x_1, x_2, \dots, x_n\}$. 因此, a 和 b 的极大似然估计为

$$\hat{a} = \min\{X_1, X_2, \dots, X_n\} = X_{(1)}, \quad \hat{b} = \max\{X_1, X_2, \dots, X_n\} = X_{(n)}.$$

同样的理由, 若用极大似然法估计例 4.5 中的 θ , 得到的结果是

$$\hat{\theta} = X_{(n)}.$$

极大似然估计

(3) Θ 为离散参数空间.

在此情况下, 为求极大似然估计, 经常考虑参数取相邻的值时, 似然函数的比值.

例 4.9 在鱼池中随机地捕捞 500 条鱼, 做上记号后再放入池中, 待充分混合后, 再捕捞 1000 条, 结果发现其中有 72 条鱼带有记号. 试问鱼池中可能有多少条鱼?

极大似然估计

解：先将问题一般化. 设池中有 N 条鱼，其中 r 条带有记号，随机地捕捞到 s 条，发现 x 条带有记号，用上述信息来估计 N .

用 X 记捕捞到的 s 条鱼中带有记号的鱼数，则有

$$P\{X = x\} = \frac{C_{N-r}^{s-x} C_r^x}{C_N^s}.$$

因此，似然函数为

$$L(N; x) = P\{X = x\},$$

考虑似然函数的比

$$g(N) = \frac{L(N; x)}{L(N-1; x)} = \frac{(N-s)(N-r)}{N(N-r-s+x)} = \frac{N^2 - (r+s)N + rs}{N^2 - (r+s)N + xN},$$

极大似然估计

当 $rs > xN$ 时, 有 $g(N) > 1$, 当 $rs < xN$ 时, 有 $g(N) < 1$. 即

$$\begin{cases} L(N; x) > L(N-1; x), & \text{当 } N < \frac{rs}{x}, \\ L(N; x) < L(N-1; x), & \text{当 } N > \frac{rs}{x}. \end{cases}$$

因此, 似然函数 $L(N; x)$ 在 $N = \frac{rs}{x}$ 附近达到极大, 注意到 N 只取正整数, 易得 N 的极大似然估计为:

$$\hat{N} = \left\lfloor \frac{rs}{x} \right\rfloor,$$

其中 $\lfloor \cdot \rfloor$ 表示下取整, 即小于该值的最大整数.

将题目中的数字代入, 得到 $\hat{N} = \left\lfloor \frac{500 \times 1000}{72} \right\rfloor = 6944$. 即鱼池中鱼的总数估计为 6944 条.

极大似然估计

(4) 如果在解 (对数) 似然方程时无法得到解析表达式, 只能采用数值方法.

例 4.10 设总体 X 服从 *Cauchy* 分布, 其概率密度函数为

$$f(x; \theta) = \frac{1}{\pi[1 + (x - \theta)^2]}, \quad -\infty < x < \infty,$$

其中 θ 为未知参数. $X_1, X_2, \dots, X_n$ 来自总体 X 的样本, 求 θ 的极大似然估计.

极大似然估计

解: Cauchy 分布的似然函数为

$$L(\theta; x) = \prod_{i=1}^n f(x_i; \theta) = \frac{1}{\pi^n} \prod_{i=1}^n \frac{1}{1 + (x_i - \theta)^2},$$

相应的对数似然函数为

$$\ln L(\theta; x) = -n \ln(\pi) - \sum_{i=1}^n \ln(1 + (x_i - \theta)^2), \quad (4.11)$$

得到对数似然方程

$$\sum_{i=1}^n \frac{x_i - \theta}{1 + (x_i - \theta)^2} = 0. \quad (4.12)$$

极大似然估计的性质

定理 7.3.3 设 $\xi \sim F(x; \theta)$, θ 为一维的未知参数, 若 $T(\xi_1, \dots, \xi_n)$ 为 θ 的充分估计量, 则似然方程

$$\frac{\partial \ln L(\xi_1, \dots, \xi_n; \theta)}{\partial \theta} = 0$$

的解一定是 $t = T(x_1, \dots, x_n)$ 的函数.

极大似然估计的性质

证明 因为 T 是 θ 的充分估计量, 由因子分解定理得

$$L(\xi_1, \dots, \xi_n; \theta) = \prod_{i=1}^n f(\xi_i, \theta) = f_T(t; \theta)h(\xi_1, \dots, \xi_n).$$

此时极大似然方程变为:

$$\frac{\partial \ln L(\xi_1, \dots, \xi_n; \theta)}{\partial \theta} = \frac{\partial \ln f_T(t, \theta)}{\partial \theta} + \frac{\partial \ln h(\xi_1, \dots, \xi_n)}{\partial \theta} = \frac{\partial \ln f_T(t, \theta)}{\partial \theta} = 0.$$

从上式的右边可以看出, θ 的极大似然估计量的解一定是 t 的函数, 也就是说它的解依赖于 $t = T(\xi_1, \dots, \xi_n)$ 的观察值.

极大似然估计的不变原理

不变原理: 设 $X \sim p(x; \theta), \theta \in \Theta$, 若 θ 的极大似然估计为 $\hat{\theta}$, 则对任意函数 $\gamma = g(\theta)$, γ 关于导出似然函数的极大似然估计为 $\hat{\gamma} = g(\hat{\theta})$.

目 录

2.1 矩估计

2.2 极大似然估计

2.3 无偏估计与一致最小方差无偏估计

2.4 完备统计量

2.5 CR不等式及有效估计

2.6 相合估计

估计量优劣的评价标准

在前两节我们介绍了矩估计和极大似然估计，并且我们知道同一个参数的矩估计和极大似然估计均不唯一。

估计量优劣的评价标准

那么当一个参数的估计量不唯一时，我们如何选取较好的估计量呢？这就涉及到估计量优劣的评价标准了，常用的评价标准有：

- 1，无偏性
- 2，相合性
- 3，渐进正态性
- 4，无偏估计的有效性
- 5，有偏估计的均方误差准则

无偏估计

定义 2.4.1 (UE) 如果 $T(\mathbf{X})$ 是未知参数 θ 的函数 $g(\theta)$ 的一个估计量, 且满足

$$E_{\theta}T(\mathbf{X}) = g(\theta), \forall \theta \in \Theta,$$

则称 T 是 $g(\theta)$ 的UE, 也称 $g(\theta)$ 是可估的(Estimable), 其中 E_{θ} 表示期望是在分布 f_{θ} 下进行的.

定义 2.4.2 (渐近UE) 如果 $T(\mathbf{X})$ 是 $g(\theta)$ 的一个有偏估计量, 且满足

$$\lim_{n \rightarrow \infty} E_{\theta}T(X_1, \dots, X_n) = g(\theta), \forall \theta \in \Theta,$$

则称 T 是 $g(\theta)$ 的渐近UE.

关于无偏估计, 我们有如下几个注解:

- 如果 $T(\mathbf{X})$ 是 $g(\theta)$ 的有偏估计, 则称 $E_{\theta}T(\mathbf{X}) - g(\theta)$ 为其偏差(bias).
- 无偏估计是从多次重复的角度而引出的一个概念, 从期望的定义不难看出, 尽管一次估计, $T(\mathbf{x})$ 的值不一定恰好等于参数真值 $g(\theta)$, 但当大量重复使用时, 其多次估计的平均值即等于参数值 $g(\theta)$.
- 一个参数的无偏估计可能不是唯一的, 也可能不存在, 也可能不合理.

刀切法

大偏差通常被视为估计的一种不足，于是，有人提出了缩小偏差的方法。下面的刀切法就是由Quenouille于1949(*JRSS B*, 11, 18-44)和1956(*Biometrika*, 43, 353-360)提出的，而正式命名则由Tukey于1958年(*Ann. Math. Statist.* 29, 614)给出。

例 2.4.1 (刀切法(Jackknife)) 设 $T(\mathbf{X})$ 是基于样本 $\mathbf{X}=(X_1, \dots, X_n)'$ 的关于参数 $g(\theta)$ 的估计量，且满足 $E_{\theta}T(\mathbf{X}) = g(\theta) + O\left(\frac{1}{n}\right)$ 。如以 $\mathbf{X}_{(-i)}$ 表示从样本中删去 X_i 后的向量，则 $T(\mathbf{X})$ 的刀切统计量定义为

$$T_J(\mathbf{X}) = nT(\mathbf{X}) - \frac{n-1}{n} \sum_{i=1}^n T(\mathbf{X}_{(-i)}). \quad (2.4.1)$$

可以证明：由(2.4.1)定义的刀切统计量具有如下性质：

$$E_{\theta}T_J(\mathbf{X}) = g(\theta) + O\left(\frac{1}{n^2}\right),$$

并且其方差不会增大。关于刀切法的详细介绍，请有兴趣的参考Efron (1982)。另外，还有一个与刀切法类似的方法—Bootstrap(自助法)，也是统计中非常有用的一种方法，具体的可见Efron and Tibshirani (1993)，和Shao and Tu (1995)。

无偏估计

前面我们讲，充分统计量具有不变性，那无偏估计具有这种不变性吗？下面的例子将回答这个问题。

例 2.4.4 对于正态总体，样本标准差 S_n 不是 σ 的UE.

解 由定理1.3.4可知，此时 $(n-1)S_n^2/\sigma^2 \sim \chi^2(n-1)$ ，故有

$$E(\sqrt{n-1}S_n/\sigma) = \int_0^\infty \frac{\sqrt{x}}{2^{(n-1)/2}\Gamma((n-1)/2)} e^{-x/2} x^{\frac{n-1}{2}-1} dx = \dots = \frac{\sqrt{2}\Gamma(n/2)}{\Gamma((n-1)/2)},$$

故

$$ES_n = \sigma \sqrt{\frac{2}{n-1}} \frac{\Gamma(n/2)}{\Gamma((n-1)/2)}, \quad (2.4.2)$$

所以，样本标准差不是 σ 的UE. □

上例告诉我们，如果 T 是 θ 的UE，则除了 g 是线性函数之外， $g(T)$ 并不是 $g(\theta)$ 的UE.

无偏估计的有效性

定义: 设 $\hat{\theta}_1 = \hat{\theta}_1(x_1, x_2, \dots, x_n)$ 与 $\hat{\theta}_2 = \hat{\theta}_2(x_1, x_2, \dots, x_n)$ 都是参数 θ 的无偏估计, 如果

$$Var(\hat{\theta}_1) \leq Var(\hat{\theta}_2), \forall \theta \in \Theta$$

且至少对一个 $\theta_0 \in \Theta$, 有严格不等号成立, 则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 有效.

无偏估计的有效性

定义: 设 $\hat{\theta}_1 = \hat{\theta}_1(x_1, x_2, \dots, x_n)$ 与 $\hat{\theta}_2 = \hat{\theta}_2(x_1, x_2, \dots, x_n)$ 都是参数 θ 的无偏估计, 如果

$$Var(\hat{\theta}_1) \leq Var(\hat{\theta}_2), \forall \theta \in \Theta$$

且至少对一个 $\theta_0 \in \Theta$, 有严格不等号成立, 则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 有效.

一致最小均方误差准则

定义 2.4.3 设 $X_1, \dots, X_n$ 是来自分布族 $\mathcal{F} = \{f(x, \theta) : \theta \in \Theta\}$ 中某一分布的样本, $g(\theta)$ 是一参数函数, 以 $\mathcal{E}(g)$ 表示用来估计 $g(\theta)$ 的某些估计量的集合, 如果存在一个 $T^* \in \mathcal{E}(g)$, 使得对任一 $T \in \mathcal{E}(g)$ 均有

$$E_{\theta}(T^* - g(\theta))^2 \leq E_{\theta}(T - g(\theta))^2, \forall \theta \in \Theta, \quad (2.4.3)$$

则称 T^* 为 $g(\theta)$ 的在 $\mathcal{E}(g)$ 中的一致最小均方误差估计, 也称满足(2.4.3)式的统计量 T^* 在均方意义下优于 T .

之所以称满足(2.4.3)式的 T^* 为最小均方误差估计, 是由于估计量 $T \in \mathcal{E}(g)$ 的均方误差(Mean Square Error, 简记为MSE)的定义为 $\text{MSE}(T) = E_{\theta}[T - g(\theta)]^2$. 当 T 是 $g(\theta)$ 的UE时, 其MSE就是它的方差.

一致最小均方误差准则

一致最小方差无偏估计 (UMVUE)

定义 2.4.4 (UMVUE) 对于参数分布族 $\mathcal{F} = \{f(x, \theta) : \theta \in \Theta\}$, $g(\theta)$ 是一可估函数, 又设 T^* 是 $g(\theta)$ 的一个 UE. 如对于 $g(\theta)$ 的任一 UE T , 均有

$$\text{Var}_{\theta}(T^*) \leq \text{Var}_{\theta}(T), \forall \theta \in \Theta, \quad (2.4.5)$$

则称 T^* 是 $g(\theta)$ 的 UMVUE.

从 UMVUE 的定义不难看出, 它是一个在无偏估计类中方差最小的估计. 但我们注意到, 虽然 UMVUE 是一个很好的估计, 但对于某些分布族或参数, 其 UMVUE 不一定存在(见本节例 2.4.5)

无偏估计的存在性

考察二项分布 $b(m, p)$, 不管样本容量 n 多大, 参数 $g(p) = 1/p$ 的无偏估计都不存在。

一致最小方差无偏估计 (UMVUE)

为了方便求取UMVUE, 我们给出如下一个定理.

定理 2.4.1 设 $T(\mathbf{X})$ 是参数分布族 $\mathcal{F} = \{f(x, \theta) : \theta \in \Theta\}$ 的一个充分统计量, $S(\mathbf{X})$ 是参数 $g(\theta)$ 的一个UE, 则

$$h(T) = E[S(\mathbf{X})|T(\mathbf{X})], \quad (2.4.6)$$

是 $g(\theta)$ 的一个UE, 且

$$\text{Var}_{\theta}(h(T)) \leq \text{Var}_{\theta}(S(\mathbf{X})), \quad \forall \theta \in \Theta, \quad (2.4.7)$$

其中等号成立的充要条件是 $P\{S(\mathbf{X}) = h(T)\} = 1$.

一致最小方差无偏估计 (UMVUE)

定理2.4.1的证明:

一致最小方差无偏估计 (UMVUE)

从定理2.4.1可以看出, 当我们找到充分统计量后, 任一个UE都可以得到改进, 并且, 为找到UMVUE, 我们仅需在基于充分统计量的无偏估计类中去找即可. 但这一定理并没有告诉我们如何才能得到UMVUE. 下一定理将回答这一问题. 为此, 我们引入如下两个无偏估计类:

$$U = \{T : E_{\theta}T = g(\theta), E_{\theta}T^2 < \infty, \forall \theta \in \Theta\}, \quad (2.4.8)$$

$$U_0 = \{T : E_{\theta}T = 0, E_{\theta}T^2 < \infty, \forall \theta \in \Theta\}, \quad (2.4.9)$$

它们分别表示 $g(\theta)$ 与0的具有二阶矩的无偏估计类.

定理 2.4.2 对于参数分布族 $\mathcal{F} = \{f(x, \theta) : \theta \in \Theta\}$, 设 $g(\theta)$ 可估. 则估计量 $T_0 \in U$ 是 $g(\theta)$ 的一个UMVUE 的充要条件是

$$E_{\theta}(\nu T_0) = 0, \forall \theta \in \Theta, \forall \nu \in U_0. \quad (2.4.10)$$

一致最小方差无偏估计 (UMVUE)

定理2.4.2的证明:

一致最小方差无偏估计 (UMVUE)

定理2.4.2的证明:

一致最小方差无偏估计 (UMVUE)

例 2.4.7 设 $X_1, \dots, X_n$ 为来自正态总体 $N(\mu, \sigma^2)$ 的iid样本, 试求 μ, σ^2 的UMVUE.

解 此时样本的联合PDF为

$$f(\mathbf{x}; \mu, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right\},$$

则任给 $\nu \in U_0$, 有

$$\int \cdots \int \nu(\mathbf{x}) \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right\} d\mathbf{x} = 0, \quad \forall \mu, \sigma^2. \quad (*)$$

上式两边关于 μ 求导, 综合(*)式后, 得到

$$\int \cdots \int \nu(\mathbf{x}) \left(\sum_{i=1}^n x_i \right) \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \right\} d\mathbf{x} = 0, \quad \forall \mu, \sigma^2, \quad (*1)$$

如取 $T(\mathbf{X}) = \sum_{i=1}^n X_i/n = \bar{X}$, 则上式即为

$$\int \cdots \int \nu(\mathbf{x}) T(\mathbf{x}) f(\mathbf{x}; \mu, \sigma^2) d\mathbf{x} = 0, \quad \forall \mu, \sigma^2.$$

又由于 $T(\mathbf{X})$ 是 μ 的UE, 故由定理2.4.2知, $T(\mathbf{X}) = \bar{X}$ 是 μ 的UMVUE.

一致最小方差无偏估计 (UMVUE)

在(*1)式两边再关于 μ 求导, 综合(*)式后, 有

$$\int \cdots \int \nu(\mathbf{x}) T^2(\mathbf{x}) f(\mathbf{x}; \mu, \sigma^2) d\mathbf{x} = 0, \quad \forall \mu, \sigma^2. \quad (*2)$$

另外, 在(*)式两边关于 σ^2 求导, 有

$$\int \cdots \int \nu(\mathbf{x}) \left[\sum_{i=1}^n (x_i - \mu)^2 \right] f(\mathbf{x}; \mu, \sigma^2) d\mathbf{x} = 0, \quad \forall \mu, \sigma^2. \quad (*3)$$

综合(*1),(*2)和(*3)式, 我们得到

$$\int \cdots \int \nu(\mathbf{x}) \left[\sum_{i=1}^n (x_i - \bar{x})^2 \right] f(\mathbf{x}; \mu, \sigma^2) d\mathbf{x} = 0, \quad \forall \mu, \sigma^2. \quad (*4)$$

又由于 $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ 是 σ^2 的UE, 则由(*4)式和定理2.4.2知, 样本方差 S_n^2 是总体方差 σ^2 的UMVUE. □

从上例可以看出, 由于修正的样本方差 S_n^{*2} 不是 σ^2 的UE, 故它肯定不是 σ^2 的UMVUE.

目 录

2.1 矩估计

2.2 极大似然估计

2.3 无偏估计与一致最小方差无偏估计

2.4 完备统计量

2.5 CR不等式及有效估计

2.6 相合估计

完备统计量

定义 2.5.1 (完备统计量) 对于参数分布族 $\mathcal{F} = \{f(x, \theta) : \theta \in \Theta\}$, 设 $T(\mathbf{X})$ 为一统计量, 如对所有满足条件

$$E_{\theta}g(T(\mathbf{X})) = 0, \forall \theta \in \Theta, \quad (2.5.1)$$

的统计量 $g(T)$, 都有

$$P_{\theta}\{g(T) = 0\} = 1, \forall \theta \in \Theta, \quad (2.5.2)$$

则称统计量 $T(\mathbf{X})$ 是完备统计量.

完备统计量

有时完备性的概念是针对分布族而言的，即下面的定义。

定义 2.5.2 (完备分布族) 对于参数分布族 $\mathcal{F} = \{f(x, \theta) : \theta \in \Theta\}$, 如果对于任一函数 $\psi(x)$, 由

$$E_{\theta}\psi(X) = 0, \forall \theta \in \Theta,$$

总可推出 $P_{\theta}\{\psi(X) = 0\} = 1$, 则称此分布族 $\mathcal{F}$ 是完备的。

由上述两个定义可以看出，完备统计量就意味着此统计量的抽样分布族是完备的。

关于完备统计量，我们不加证明地引入如下两个结论(证明可参见Shao (2003)):

- 如果 T 是完备的, $S = \psi(T)$, 且 ψ 可测, 则 S 也是完备的;
- 充分完备统计量是极小充分统计量, 但反之不成立.

完备统计量

例 2.5.1 考虑二项分布族 $\{B(n, p) : 0 < p < 1\}$ 的完备性.

解 设函数 $\psi(x)$ 满足

$$E_p \psi(X) = \sum_{x=0}^n \psi(x) \binom{n}{x} p^x (1-p)^{n-x} = 0, \forall 0 < p < 1.$$

令 $\theta = p/(1-p)$, 则 $\theta > 0$, 且上式可以写成

$$\sum_{x=0}^n \psi(x) \binom{n}{x} \theta^x = 0, \forall \theta > 0.$$

由于上式左边是 θ 的一个 n 次多项式, 故有 $\psi(x) = 0, x = 0, 1, \dots, n$, 故二项分布是完备的. □

完备统计量的应用

定理 2.5.1 设 $X_1, \dots, X_n$ 是来自参数分布族 $\{f(x, \theta) : \theta \in \Theta\}$ 的iid样本, T 是 θ 的充分完备统计量. 如果 θ 可估, 且 $S(\mathbf{X})$ 是 θ 的一个UE, 则 $S_0(T) = E[S(\mathbf{X})|T]$ 是 θ 的唯一的UMVUE.

完备统计量的应用

例 2.5.4 考虑Bernoulli总体 $b(1, p)$ 中参数 p 的UMVUE.

解 对于此问题, 由例1.4.1可知, 统计量 $T_n = \sum_{i=1}^n X_i$ 是充分的, 而由例2.5.1可知, 二项分布族是完备的, 而 $T_n \sim B(n, p)$, 故知统计量 T_n 是充分完备的. 又由于 $\bar{X} = T_n/n$ 是 p 的UE, 且是充分完备统计量 T_n 的函数, 故由定理2.5.1知, $\bar{X}$ 是 p 的UMVUE. $\square$

指数分布族的充分完备性

定理 2.5.2 设 $\{f(x, \eta) : \eta \in \Xi\}$ 是 k 个参数的标准指数型分布族, 即

$$f(x, \eta) = c(\eta) \exp \left\{ \sum_{j=1}^k \eta_j T_j(x) \right\} h(x), \quad (2.5.3)$$

其中 $\eta = (\eta_1, \dots, \eta_k)' \in \Xi$, 这里

$$\Xi = \left\{ (\eta_1, \dots, \eta_k) : 0 < \int_{\mathcal{X}} \exp \left\{ \sum_{j=1}^k \eta_j T_j(x) \right\} h(x) dx < \infty \right\} \quad (2.5.4)$$

为自然参数空间. 则

- 统计量 $T(\mathbf{X}) = (T_1, \dots, T_k)$ 是充分的;
- 如果 Ξ 包含一个开集 (此时, 称此分布族为满秩的), 则 $T(\mathbf{X}) = (T_1, \dots, T_k)$ 是完备的.

指数分布族的充分完备性

例 2.5.7 考虑Poisson分布族 $P(\lambda)$ 中参数 λ 的UMVUE.

解 由Poisson分布的可加性知, 统计量 $T_n = \sum_{i=1}^n X_i \sim P(n\lambda)$, 且其CDF为

$$P\{T = t\} = \frac{(n\lambda)^t}{t!} e^{-n\lambda} = e^{-n\lambda} \exp\{t \ln(n\lambda)\} / t!, \quad t = 0, 1, \dots$$

如令 $\eta = \ln(n\lambda)$, 则知其自然参数空间为 $\Omega = R$, 它有内点, 故由定理2.5.2知, 统计量 $T = \sum_{i=1}^n X_i$ 是充分完备的. 另外, 由于 $\bar{X} = T/n$ 是 λ 的UE, 故它是 λ 的UMVUE. $\square$

指数分布族的充分完备性

例 2.5.8 求Gamma分布族 $\Gamma(\alpha, \lambda)$ 的充分完备统计量.

解 设 $X_1, \dots, X_n$ 为来自 $\Gamma(\alpha, \lambda)$ 的iid样本, 由其联合PDF为

$$f(\mathbf{x}; \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} = \frac{\lambda^\alpha}{\Gamma(\alpha)} \exp \left\{ (\alpha - 1) \sum_{i=1}^n \ln x_i - \lambda \sum_{i=1}^n x_i \right\},$$

于是, 由定理2.5.2可知, 统计量 $(\sum_{i=1}^n X_i, \sum_{i=1}^n \ln X_i)$ 是 (α, λ) 的充分完备统计量. □

指数分布族的充分完备性

例 7.3.10 设 $\xi \sim N(a, \sigma)$, $\xi_1, \dots, \xi_n$ 为其子样. 试证, $\bar{\xi}$ 及 S^2 是 a 及 σ^2 的充分完备估计量.

指数分布族的充分完备性

证明 $\xi_1, \dots, \xi_n$ 的联合密度函数为

$$\begin{aligned} f(x_1, \dots, x_n) &= \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - a)^2 \right\} \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \cdot e^{-\frac{na^2}{2\sigma^2}} \cdot \exp \left\{ -\frac{n}{2\sigma^2} \left(\frac{1}{n} \sum_{i=1}^n x_i^2 \right) + \frac{na}{\sigma^2} \bar{x} \right\} \end{aligned}$$

取

$$\theta = (a, \sigma^2), \quad C(\theta) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{na^2}{2\sigma^2}},$$

$$T(X) = \left(\bar{X}, \frac{1}{n} \sum_{i=1}^n x_i^2 \right), \quad b(\theta) = \left(\frac{na}{\sigma^2}, -\frac{n}{2\sigma^2} \right), \quad h(X) = 1,$$

指数分布族的充分完备性

$$\left(\text{即 } T(X) = (T_1(X), T_2(X)), T_1(X) = \bar{X}, T_2(X) = \frac{1}{n} \sum_{i=1}^n x_i^2; b(\theta) = (b_1(\theta), b_2(\theta)), \right.$$

$$\left. b_1(\theta) = \frac{na}{\sigma^2}, b_2(\theta) = -\frac{n}{2\sigma^2} \right)$$

故

$$f(x_1, \dots, x_n) = C(\theta) \exp\{b_1(\theta)T_1(X) + b_2(\theta)T_2(X)\}h(X).$$

注意到 $\Omega = (-\infty, +\infty) \times [0, +\infty)$, $b(\theta) = \left(\frac{na}{\sigma^2}, -\frac{n}{2\sigma^2}\right)$, 满足定理 7.3.8 的条件, 故 $T(X)$ 是 θ 的充分完备估计量, 即 $\bar{\xi}, S^2$ 是 a, σ^2 的充分完备估计量. 同理 $\bar{\xi}, S^{*2} = \frac{n}{n-1} S^2$ 是 a, σ^2 的无偏估计量, 因而是 a, σ^2 的最优无偏估计量.

通过充分完备统计量寻找UMVUE

通过充分完备统计量寻找UMVUE

目 录

2.1 矩估计

2.2 极大似然估计

2.3 无偏估计与一致最小方差无偏估计

2.4 完备统计量

2.5 **CR**不等式及有效估计

2.6 相合估计

C-R不等式

对于一个估计量来说，在给定无偏性的条件下，我们总是希望该估计量的方差越小越好，然而一个无偏估计量的方差有没有一个下界呢？C-R不等式给我们做了一个回答。

瑞典统计学家克拉梅 和 印度统计学家劳 分别在1945年和1946年对单参数正则分布族证明了一个重要的不等式，后人称为Cramer-Rao不等式，简称C-R不等式。

这个不等式给出了可估参数的无偏估计的方差下界，这个下界与三个量有关：

- 1，样本量 n ;
- 2，Fisher信息量 $I(\theta)$
- 3，可估参数 $g(\theta)$ 的变化率

正则分布族与Fisher信息量

定义 2.6.1 (正则分布) 如果单参数分布族 $\mathcal{F} = \{f(x, \theta) : \theta \in \Theta\}$ 具有如下五个条件:

- (1) 参数空间 Θ 是直线上的开区间(有限、无限或半无限);
- (2) 导数 $\frac{\partial f(x, \theta)}{\partial \theta}$ 存在($\forall \theta \in \Theta$);
- (3) 支撑集与 θ 无关;
- (4) 其PDF $f(x, \theta)$ 的积分与微分运算可以互换, 即

$$\frac{d}{d\theta} \int_{-\infty}^{+\infty} f(x, \theta) dx = \int_{-\infty}^{+\infty} \frac{\partial}{\partial \theta} f(x, \theta) dx.$$

(5)

$$I(\theta) = E_{\theta} \left(\frac{\partial}{\partial \theta} \ln f(X, \theta) \right)^2 \text{ 存在, 且 } I(\theta) > 0, \quad (2.6.1)$$

则称此分布族为C-R正则分布族, 其中条件(1)–(5)也称为正则条件, $I(\theta)$ 称为该分布族的Fisher 信息量(Information).

C-R不等式

定理 2.6.2 (信息不等式) 设分布族 $\mathcal{F} = \{f(x, \theta) : \theta \in \Theta\}$ 是正则的, 可估函数 $g(\theta)$ 在 Θ 上可微. 又设 $X_1, \dots, X_n$ 是 n 个来自此分布族的iid样本, $T(\mathbf{X})$ 是 $g(\theta)$ 的一个UE, 且满足积分与微分号可互换的条件, 即

$$\frac{d}{d\theta} \int T(\mathbf{x})f(\mathbf{x}, \theta)d\mathbf{x} = \int T(\mathbf{x})\frac{\partial}{\partial\theta}f(\mathbf{x}, \theta)d\mathbf{x},$$

则有

$$\text{Var}_{\theta}(T(\mathbf{X})) \geq \frac{(g'(\theta))^2}{nI(\theta)}, \quad \forall \theta \in \Theta, \quad (2.6.10)$$

其中 $I(\theta)$ 为 $\mathcal{F}$ 的Fisher信息量. 如果上式等号成立, 则存在 $c(\theta) \neq 0$, 使得

$$\frac{\partial \ln f(\mathbf{X}, \theta)}{\partial \theta} = c(\theta)(T(\mathbf{X}) - g(\theta)) \quad (2.6.11)$$

以概率1成立.

C-R不等式的证明

C-R不等式的证明:

C-R不等式的证明

C-R不等式的证明:

C-R不等式

注 2.6.1 我们称不等式(2.6.10)为信息不等式, 也称C-R不等式.

注 2.6.2 在上一定理的证明过程中, 我们始终假设样本是iid的. 如果样本不是独立的, 上述结论仍成立, 只是要把(2.6.10)中的 $nI(\theta)$ 改为 $E_{\theta} \left(\frac{\partial \ln f(\mathbf{X}, \theta)}{\partial \theta} \right)^2$.

注 2.6.3 从(2.6.10)可以看出, 信息不等式与Fisher信息量有着密切的关系. 从直观上看, 一个带有较小方差的估计表明这个估计值有更大机会出现在 $g(\theta)$ 附近. 为了考察 $I(\theta)$ 的含义, 我们假设信息不等式的下界可以达到, 且 $g(\theta) = \theta$, 则此时UE的最小方差为 $\frac{1}{nI(\theta)}$. 这说明, $nI(\theta)$ 越大, 参数 θ 可以越精确地估计, 其中 n 是样本容量, 这也说明, 如以估计量的方差的倒数作为估计量精度的指标, 则精度与 n 成正比, 而 $I(\theta)$ 则反映着总体分布的性质. $I(\theta)$ 越大, 说明总体本身提供的信息量越多.

注 2.6.4 从定理的证明不难看出, 对于正则分布族, 我们有如下结论:

$$E_{\theta} \left[\frac{\partial}{\partial \theta} \ln f(X, \theta) \right] = 0, \quad I(\theta) = \text{Var}_{\theta} \left[\frac{\partial}{\partial \theta} \ln f(X, \theta) \right] = -E_{\theta} \left[\frac{\partial^2}{\partial \theta^2} \ln f(X, \theta) \right]. \quad (2.6.12)$$

CR不等式的应用

例 2.6.8 考虑均匀分布 $U(0, \theta)$ 中参数 θ 的估计问题, 其中 $\theta > 0$.

解 此时总体PDF为

$$f(x, \theta) = \begin{cases} \frac{1}{\theta}, & 0 < x < \theta, \\ 0, & \text{其它}. \end{cases}$$

我们注意到, 对于固定的 $x > 0$, 由于 $f(x, \theta)$ 作为 θ 的函数在 $\theta = x$ 处不连续, 故它关于 θ 的导数不存在. 于是, 此均匀分布族不是正则的, 前面讲的C-R下界也不成立. 但如果我们不顾条件, 而要利用定理2.6.2的结论, 则会导致错误. 事实上, 由于

$$\frac{\partial \ln f(x, \theta)}{\partial \theta} = -\frac{1}{\theta}, \quad \forall \theta > 0,$$

故 $I(\theta) = \frac{1}{\theta^2}$, 从而C-R下界为 θ^2/n .

但是, 从例2.4.9可知, θ 的UMVUE为 $\frac{n+1}{n}X_{(n)}$. 计算一下可知,

$$\text{Var} \left(\frac{n+1}{n}X_{(n)} \right) = \frac{\theta^2}{n(n+2)} < \frac{\theta^2}{n} \text{---C-R下界}.$$

C-R不等式与UMVUE

例 2.6.4 设 $X_1, \dots, X_n$ 为来自Poisson分布 $P(\lambda)$ 的iid样本, 考虑 λ 的UMVUE.

解 由例2.6.1可知, $P(\lambda)$ 是正则的, 且其Fisher信息量为 $I(\lambda) = \frac{1}{\lambda}$, 于是 λ 的UE的方差的C-R下界为 $\frac{\lambda}{n}$. 又由于 $\sum_{i=1}^n X_i \sim P(n\lambda)$, 则知 $\text{Var} \bar{X} = \frac{\lambda}{n}$, 达到了C-R下界, 故知样本均值 $\bar{X}$ 是 λ 的UMVUE. □

C-R不等式与UMVUE

例 2.6.5 设 $X_1, \dots, X_n$ 是来自指数分布 $E(\lambda)$ 的iid样本, 考虑 $1/\lambda$ 的UMVUE.

解 容易验证此分布族是正则的, 且由于

$$f(x, \lambda) = \lambda e^{-\lambda x}, \ln f(x, \lambda) = \ln \lambda - \lambda x, \frac{\partial \ln f(x, \lambda)}{\partial \lambda} = \frac{1}{\lambda} - x,$$

则

$$I(\lambda) = E_{\lambda} \left(\frac{\partial \ln f(X, \lambda)}{\partial \lambda} \right)^2 = E_{\lambda} \left(\frac{1}{\lambda} - X \right)^2 = \frac{1}{\lambda^2}.$$

于是, $1/\lambda$ 的UE的方差的C-R下界为

$$\frac{(g'(\lambda))^2}{nI(\lambda)} = \frac{1}{n\lambda^2}.$$

因为 $X_i \sim E(\lambda) = \Gamma(1, \lambda)$, 而 Γ 分布关于形状参数具有可加性, 所以, $\sum_{i=1}^n X_i \sim \Gamma(n, \lambda)$. 于是, $\text{Var}(\sum_{i=1}^n X_i) = \frac{n}{\lambda^2}$, 即 $\text{Var}\bar{X} = \frac{1}{n\lambda^2}$ 达到了C-R下界, 故知 $\bar{X}$ 是 $1/\lambda$ 的UMVUE, 这与例2.4.7的结果一样. □

有效估计

定义 2.6.2 (有效估计及效率) 设 $T(\mathbf{X})$ 是 $g(\theta)$ 的一个UE, 则比值

$$e_n = \frac{(g'(\theta))^2/nI(\theta)}{\text{Var}_\theta T(\mathbf{X})} \quad (2.6.14)$$

为 $T(\mathbf{X})$ 的效率. 如果 $e_n = 1$, 则称 $T(\mathbf{X})$ 为 $g(\theta)$ 的有效估计. 如 $\lim_{n \rightarrow \infty} e_n = 1$, 则称 $T(\mathbf{X})$ 为 $g(\theta)$ 的渐近有效估计.

目 录

2.1 矩估计

2.2 极大似然估计

2.3 无偏估计与一致最小方差无偏估计

2.4 完备统计量

2.5 CR不等式及有效估计

2.6 相合估计

相合估计

前几节的内容大都是固定样本量 n 而讨论估计的性质。实际上，估计的大样本性质也是人们非常关心的一个问题。

接下来，我们将介绍估计量的大样本性质：**相合估计**和**相合渐近正态估计**。大样本性质在进行理论研究时，具有重要意义。

相合估计

定义 2.7.1 设统计量 T_n 是总体参数 $g(\theta)$ 的估计量.

- 如果当 $n \rightarrow \infty$ 时, T_n 依概率收敛于 $g(\theta)$, 即 $\forall \theta \in \Theta$ 及 $\epsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P\{|T_n - g(\theta)| \geq \epsilon\} = 0, \quad (2.7.1)$$

则称 T_n 是 $g(\theta)$ 的(弱)相合估计.

- 如果当 $n \rightarrow \infty$ 时, T_n 以概率1收敛于 $g(\theta)$, 即 $\forall \theta \in \Theta$, 有

$$P\{\lim_{n \rightarrow \infty} T_n = g(\theta)\} = 1, \quad (2.7.2)$$

则称 T_n 是 $g(\theta)$ 的强相合估计.

- 如果当 $n \rightarrow \infty$ 时, T_n 依 r 阶矩收敛于 $g(\theta)$, 即 $\forall \theta \in \Theta$, 有

$$\lim_{n \rightarrow \infty} E_{\theta}|T_n - g(\theta)|^r = 0, \quad (2.7.3)$$

则称 T_n 是 $g(\theta)$ 的 r 阶矩相合估计. 当 $r = 2$ 时, 称为均方相合估计.

由概率论知识知道, 强相合 $\Rightarrow$ 弱相合, r 阶矩相合 $\Rightarrow$ 弱相合, 反之不成立, 且强相合与 r 阶相合之间没有包含关系.

相合估计

例 2.7.1 设 $X_1, \dots, X_n$ 是来自 $b(1, p)$ 的iid样本, 则由大数定律知

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} p, \forall 0 < p < 1,$$

于是, $\bar{X}$ 是 p 的相合估计.

□

相合估计

由概率论知识, 容易得到如下一些结论.

定理 2.7.1 设 $X_1, \dots, X_n$ 是来自分布族 $\{f(x, \theta) : \theta \in \Theta\}$ 的iid样本, 且 $E|X_1|^p < \infty$ (p 为正整数), 则样本的 $k(1 \leq k \leq p)$ 阶原点矩是总体 k 阶原点矩的相合估计, 即

$$a_k = \frac{1}{n} \sum_{i=1}^n X_i^k \xrightarrow{P} \mu_k = EX^k.$$

定理 2.7.2 如果 T_n 是 $g(\theta)$ 的相合估计, c_n, d_n 是两个常数列, 且 $\lim_n c_n = 0, \lim_n d_n = 1$, 则 $d_n T_n + c_n$ 也是 $g(\theta)$ 的相合估计.

定理 2.7.3 如果 T_n 是 $g(\theta)$ 的渐近无偏估计, 且满足

$$\lim_{n \rightarrow \infty} \text{Var}_{\theta} T_n = 0, \forall \theta \in \Theta,$$

则 T_n 既是 $g(\theta)$ 的相合估计, 也是均方相合估计.

定理 2.7.4 如果 T_n 是 θ 的相合估计, $g(x)$ 在点 $x = \theta$ 处连续, 则 $g(T_n)$ 也是 $g(\theta)$ 的相合估计.

由上述定理很容易知道, 样本方差是总体方差的相合估计, 多数矩估计也是相应参数的相合估计. 关于矩估计的相合性, 请参见Serfling (1980) §2.2.

样本分位数的相合性

§2.7.2 样本分位数的相合性

由于分位数在数理统计中起着非常重要的作用，而我们在第一章中讲述了样本 p 分位数的概念，下面看一看其大样本性质如何. 为此，我们先引进总体 p 位数的概念.

定义 2.7.2 (总体 p 分位数) 设总体 X 的分布函数为 $F(x)$ ，对于给定的 $p \in (0, 1)$ ，称满足

$$\xi_p = \inf\{x : F(x) \geq p\}, \quad (2.7.4)$$

的 ξ_p 为该总体的 p 分位数. 特别地， $\xi_{0.5}$ 称为该总体的中位数.

定理 2.7.5 (样本分位数的Bahadur表示) 以 $F_n(x)$ 表示经验分布函数， ξ_p 表示总体的 p 分位数， $f(x)$ 表示总体的PDF. 如果 $f(\xi_p) > 0$ 且 $f(x)$ 在 ξ_p 点连续，则当 $n \rightarrow \infty$ 时，有

$$\sqrt{n} \left(m_{n,p} - \xi_p + \frac{F_n(\xi_p) - p}{f(\xi_p)} \right) \xrightarrow{P} 0. \quad (2.7.5)$$

此定理的证明请参见陈希孺等(1989).

极大似然估计的相合性

§2.7.3 极大似然估计的相合性

本小节将讨论MLE的相合性, 为此假设 $X_1, \dots, X_n$ 为来自 $f(x; \theta)$ 的IID样本, 为简单起见, 我们只考虑单参数的情况, 设参数空间 Θ 是一个开区间, 对数似然函数为 $l(\theta; x) = \sum_{i=1}^n \ln f(x_i; \theta)$.

定理 2.7.6 如 $\ln f(x; \theta)$ 在 Θ 上可微, 并设 $f(x; \theta)$ 是可识别的, 即 $\forall \theta \neq \theta', \{x; f(x; \theta) \neq f(x; \theta')\}$ 不是零测集. 则似然方程在 $n \rightarrow \infty$ 时以概率1有解, 且此解关于 θ 是相合的.

相合渐近正态估计

定义 2.7.3 (CAN) 设 T_n 是参数 $g(\theta)$ 的相合估计量, 如存在与样本容量 n 有关的定义于参数空间 Θ 上的函数 $\mu_n(\theta), \sigma_n(\theta)$, 且 $\sigma_n(\theta) > 0$, 使得当 $n \rightarrow \infty$ 时有

$$\frac{T_n - \mu_n(\theta)}{\sigma_n(\theta)} \xrightarrow{\mathcal{L}} N(0, 1), \quad (2.7.6)$$

则称 T_n 为 $g(\theta)$ 的CAN估计, 也称 T_n 渐近正态 $N(\mu_n, \sigma_n^2)$, 记为 $T_n \sim AN(\mu_n, \sigma_n^2)$.

相合渐近正态估计

一、样本分位数的相合渐近正态性

定理 2.7.7 设 ξ_p 表示总体的 p 分位数, $f(x)$ 表示总体的PDF. 如果 $f(\xi_p) > 0$ 且 $f(x)$ 在 ξ_p 点连续, 则当 $n \rightarrow \infty$ 时, 有

$$\sqrt{n}(m_{n,p} - \xi_p) \xrightarrow{\mathcal{L}} N(0, p(1-p)/f^2(\xi_p)). \quad (2.7.7)$$

相合渐近正态估计

二、矩估计的相合渐近正态性

设 $X_1, \dots, X_n$ 为来自某参数分布族 $\{f(x, \theta) : \theta \in \Theta\}$ 的iid样本, 且待估参数 $g(\theta)$ 可以表示为总体 k 阶原点矩 μ_k 的函数, 即

$$g(\theta) = G(\mu_1, \dots, \mu_k).$$

如用样本 k 阶原点矩估计总体 k 阶原点矩, 则记得到的关于 $g(\theta)$ 的估计为

$$\hat{g}(\mathbf{X}) = G(a_1, \dots, a_k).$$

再设总体的 $2k$ 阶原点矩存在且函数 G 关于各变量均一阶连续可微, 令

$$B = (b_{ij})_{k \times k} = (\mu_{i+j} - \mu_i \mu_j)_{k \times k}, \quad d_i = \frac{\partial G}{\partial \mu_i}, \quad i = 1, \dots, k,$$

并且 $\mathbf{d} = (d_1, \dots, d_k)'$, $\sigma^2 = \mathbf{d}' B \mathbf{d}$.

则在上述假设之下, 我们有如下的结论.

定理 2.7.8 在上面条件下, 当 $n \rightarrow \infty$ 时, 有

$$\sqrt{n} (\hat{g}(\mathbf{X}) - G(\mu_1, \dots, \mu_k)) \xrightarrow{\mathcal{L}} N(0, \sigma^2). \quad (2.7.8)$$

相合渐近正态估计

三、极大似然估计的相合渐近正态性

为了简单明了, 下面我们仅给出了总体参数为一维的情形.

定理 2.7.9 设参数空间 Θ 是开区间, 总体PDF $f(x; \theta)$ 满足

- (1) 在参数真值 θ_0 的邻域内, $\partial^i \ln f / \partial \theta^i$, $i = 1, 2, 3$ 存在;
- (2) 在参数真值 θ_0 的邻域内, $|\partial^3 \ln f / \partial \theta^3| \leq H(x)$, 且 $EH(X) < \infty$;
- (3) 在参数真值 θ_0 处,

$$E_{\theta_0} \left[\frac{f'(X, \theta_0)}{f(X, \theta_0)} \right] = 0, \quad E_{\theta_0} \left[\frac{f''(X, \theta_0)}{f(X, \theta_0)} \right] = 0, \quad I(\theta_0) = E_{\theta_0} \left[\frac{f'(X, \theta_0)}{f(X, \theta_0)} \right]^2 > 0.$$

记 $\hat{\theta}_n$ 为当 $n \rightarrow \infty$ 时似然方程的相合解, 则

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} N(0, I^{-1}(\theta_0)). \quad (2.7.9)$$

我们在这里

□ 问题答疑：<http://www.xxwenda.com/>

■ 可邀请老师或者其他回答问题

联系我们

小象学院：互联网新技术在线教育领航者

- 微信公众号：小象
- 新浪微博：ChinaHadoop

