



第七章 参数估计

• 中山大学 • 计算机学院 • 郑培嘉 • Email: zhpj@mail.sysu.edu.cn

目录

1. 点估计
2. * 基于截尾样本的最大似然估计
3. 估计量的评选标准
4. 区间估计
5. 正态总体均值与方差的区间估计
6. (0-1)分布参数的区间估计
7. 单侧置信区间



参数估计问题是利用从总体抽样得到的信息来估计总体的某些参数或者参数的某些函数.

例：在某炸药制造厂，一天中发生着火现象的次数 X 是一个随机变量, 假设它服从以 $\lambda > 0$ 为参数的泊松分布, 参数 λ 为未知, 设有以下的样本值, 试估计参数 λ .

着火次数 k	0	1	2	3	4	5	6
发生 k 次着火的天数 n_k	75	90	54	22	6	2	1

在参数估计问题中, 假定总体分布形式已知, 未知的仅仅是一个或几个参数.

解：由于 $X \sim \pi(\lambda)$, 故有 $\lambda = E(X)$. 我们自然想到用样本均值来估计总体的均值 $E(X)$, 由已知数据计算得到

$$\bar{x} = \frac{\sum_{k=0}^6 k n_k}{\sum_{k=0}^6 n_k}$$

$$\begin{aligned} &= \frac{1}{250} [0 \times 75 + 1 \times 90 + 2 \times 54 + 3 \times 22 + 4 \times 6 + 5 \times 2 + 6 \times 1] \\ &= 1.22 \end{aligned}$$

即 $E(X) = \lambda$ 的估计为 1.22



参数估计问题的定义

设有一个统计总体，总体的分布函数为 $F(x, \theta)$ ，其中 θ 为未知参数(θ 可以是向量)。现从该总体抽样，得样本

$$X_1, X_2, \dots, X_n$$

要依据该样本对参数 θ ，作出估计，或估计 θ 的某个已知函数 $g(\theta)$ 。

这类问题称为参数估计。



参数估计 { 点估计
 { 区间估计



例如我们要估计某队男生的平均身高. 假定身高服从正态分布 $N(\mu, 0.1^2)$.

现从该总体选取容量为5的样本, 我们的任务是要根据选出的样本 (5个数) 求出总体均值 μ 的估计. 而全部信息就由这5个数组成.

设这5个数是:

1.65 1.67 1.68 1.78 1.69

估计 μ 为1.68, 这是点估计.

估计 μ 在区间 $[1.57, 1.84]$ 内, 这是区间估计.



寻求估计量的方法

1. 极大似然法
2. 矩估计法
3. 最小二乘法
4. 贝叶斯方法

这里我们主要介绍前面两种方法 .





1. 点估计



点估计

点估计问题的一般提法如下：设总体 X 的分布函数 $F(x; \theta)$ 的形式为已知， θ 是待估参数。 X_1, X_2, \dots, X_n 是 X 的一个样本， x_1, x_2, \dots, x_n 是相应的一个样本值。点估计问题就是要构造一个适当的统计量 $\hat{\theta} = X_1, X_2, \dots, X_n$ ，用它的观察值 $\hat{\theta} = x_1, x_2, \dots, x_n$ 作为未知参数 θ 的近似值。我们称 $\hat{\theta} = X_1, X_2, \dots, X_n$ 为 θ 的估计量，称 $\hat{\theta} = x_1, x_2, \dots, x_n$ 为 θ 的估计值。在不致混淆的情况下统称估计量和估计值为估计，并都简记为 $\hat{\theta}$ 。



点估计

下面介绍两种常用的构造估计量的方法：矩估计法和最大似然估计法。

最大似然法

它是在总体分布已知条件下使用的一种参数估计方法。

最大似然估计法是求点估计的一种方法。它最早是由德国数学家高斯于1821年所提出，后来为英国统计学家费希尔在1912年重新提出并做了进一步的研究。这是目前仍然得到最广泛应用的一种方法。它是建立在极大似然原理的基础上的一个统计方法。



点估计

最大似然法原理直观想法：“概率最大的事件最可能出现”。
例如有一个事件，若知道它出现的概率只能是0.01或0.99，而在一次观测中，此事件出现，此时自然会说它的概率应为0.99。
因此，参数估计的极大似然法是要选取这样的值来作为参数的估计值，使得当参数取这一数值时，观测结果出现的可能性为最大。



点估计

思想方法：使得事件的出现有较大的概率

例：有两外形相同的箱子, 各装100个球

一箱 99个白球 1 个红球

一箱 1 个白球 99个红球

现从两箱中任取一箱, 并从箱中任取一球,

结果所取得的球是白球.

问：所取的球来自哪一箱？

答：第一箱.



点估计

最大似然估计法:

似然函数的定义

1) 设总体 X 属离散型。

设分布律 $P\{X = k\} = p(x; \theta)$, θ 为待估参数, $\theta \in \Theta$,

(其中 Θ 是 θ 可能的取值范围)

X_1, X_2, \dots, X_n 是来自总体 X 的样本,

则 X_1, X_2, \dots, X_n 的联合分布律为 $\prod_{i=1}^n p(x_i; \theta)$



点估计

又设 x_1, x_2, \dots, x_n 为相应于样本 X_1, X_2, \dots, X_n 的一个样本值.
则样本 X_1, X_2, \dots, X_n 取到观察值 x_1, x_2, \dots, x_n 的概率,
即事件 $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ 发生的概率为

$$L(\theta) = L(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n p(x_i; \theta), \quad \theta \in \Theta,$$

$L(\theta)$ 称为样本的似然函数.



点估计

选取使似然函数 $L(\theta)$

取得最大值的 $\hat{\theta}$ 作为未知参数 θ 的估计值,

即 $L(x_1, x_2, \dots, x_n; \hat{\theta}) = \max_{\theta \in \Theta} L(x_1, x_2, \dots, x_n; \theta)$. (其中 Θ 是 θ 可能的取值范围)

这样得到的 $\hat{\theta}$ 与样本值 x_1, x_2, \dots, x_n 有关, 记为

$$\hat{\theta}(x_1, x_2, \dots, x_n),$$

称为参数 θ 的最大似然估计值,

$$\hat{\theta}(X_1, X_2, \dots, X_n)$$

称为参数 θ 的最大似然估计量.



点估计

似然函数的定义

2) 设总体 X 属连续型。

设概率密度为 $f(x; \theta)$, θ 为待估参数, $\theta \in \Theta$,

(其中 Θ 是 θ 可能的取值范围)

X_1, X_2, \dots, X_n 是来自总体 X 的样本,

则 X_1, X_2, \dots, X_n 的联合密度为 $\prod_{i=1}^n f(x_i; \theta)$.



点估计

又设 x_1, x_2, \dots, x_n 为相应于样本 X_1, X_2, \dots, X_n 的一个样本值. 则随机点 (X_1, X_2, \dots, X_n) 落在点 (x_1, x_2, \dots, x_n) 的邻域(边长分别为 dx_1, dx_2, \dots, dx_n 的 n 维立方体)内的概率近似地为

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

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

$L(\theta)$ 称为样本的似然函数.

若 $L(x_1, x_2, \dots, x_n; \hat{\theta}) = \max_{\theta \in \Theta} L(x_1, x_2, \dots, x_n; \theta).$

$$\hat{\theta}(x_1, x_2, \dots, x_n),$$

称为参数 θ 的最大似然估计值,

$$\hat{\theta}(X_1, X_2, \dots, X_n)$$

称为参数 θ 的最大似然估计量.



点估计

两点说明:

1、求似然函数 $L(\theta)$ 的最大值点，可以应用微积分中的技巧。由于 $\ln(x)$ 是 x 的增函数， $\ln L(\theta)$ 与 $L(\theta)$ 在 θ 的同一值处达到它的最大值，假定 θ 是一实数，且 $\ln L(\theta)$ 是 θ 的一个可微函数。通过求解方程：

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

可以得到 θ 的MLE.

若 θ 是向量，上述方程必须用方程组代替.

2、用上述求导方法求参数的MLE有时行不通，这时要用最大

似然原则来求



点估计

求最大似然估计量的步骤:

(一) 写出似然函数

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

或

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

(二) 取对数

$$\ln L(\theta) = \sum_{i=1}^n \ln p(x_i; \theta) \quad \text{或} \quad \ln L(\theta) = \sum_{i=1}^n \ln f(x_i; \theta);$$



点估计

求最大似然估计量的步骤(Cont.):

(三) 对 θ 求导 $\frac{d \ln L(\theta)}{d\theta}$, 并令 $\frac{d \ln L(\theta)}{d\theta} = 0$, 解方程即得未知参数 θ 的最大似然估计值 $\hat{\theta}$.

对数似
然方程

最大似然估计法也适用于分布中含有多个未知参数的情况。
此时只需令

$$\frac{\partial}{\partial \theta_i} \ln L = 0, \quad i = 1, 2, \dots, k.$$

对数似然
方程组

解出由 k 个方程组成的方程组, 即可得各未知参数 θ_i ($i = 1, 2, \dots, k$) 的最大似然估计值 $\hat{\theta}_i$.



点估计

例： 设 $X \sim B(1, p)$, X_1, X_2, \dots, X_n 是来自 X 的一个样本, 求 p 的最大似然估计量.

解： 设 x_1, x_2, \dots, x_n 为相应于样本 X_1, X_2, \dots, X_n 的一个样本值, X 的分布律为

$$P\{X = x\} = p^x(1 - p)^{1-x}, \quad x = 0, 1,$$

似然函数：

$$\begin{aligned} L(p) &= \prod_{i=1}^n p^{x_i}(1 - p)^{1-x_i} \\ &= p^{\sum_{i=1}^n x_i} (1 - p)^{n - \sum_{i=1}^n x_i}, \end{aligned}$$



点估计

$$\ln L(p) = \left(\sum_{i=1}^n x_i \right) \ln p + \left(n - \sum_{i=1}^n x_i \right) \ln(1 - p),$$

令

$$\frac{d}{dp} \ln L(p) = \frac{\sum_{i=1}^n x_i}{p} - \frac{n - \sum_{i=1}^n x_i}{1 - p} = 0,$$

解得 p 的最大似然估计值

$$p = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}.$$



点估计

例： 设总体 $X \sim N(\mu, \sigma^2)$, μ, σ^2 为未知参数, x_1, x_2, \dots, x_n 是来自 X 的一个样本值, 求 μ 和 σ^2 的最大似然估计量.

解： X 的概率密度为

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

X 的似然函数:

$$\begin{aligned} L(\mu, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \\ &= (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right] \end{aligned}$$



点估计

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

$$\text{令} \begin{cases} \frac{\partial}{\partial \mu} \ln L(\mu, \sigma^2) = 0, \\ \frac{\partial}{\partial \sigma^2} \ln L(\mu, \sigma^2) = 0, \end{cases}$$

$$\begin{cases} \frac{1}{\sigma^2} \left[\sum_{i=1}^n x_i - n\mu \right] = 0 , \\ -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n (x_i - \mu)^2 = 0 , \end{cases}$$



点估计

由 $\frac{1}{\sigma^2} [\sum_{i=1}^n x_i - n\mu] = 0$ 解得

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

由 $-\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n (x_i - \mu)^2 = 0$ 解得

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$



点估计

例： 设总体 X 在 $[0, \theta]$ 上服从均匀分布, 其中 θ 未知, x_1, x_2, \dots, x_n 是来自总体 X 的一个样本值, 求 θ 的最大似然估计值.

解： X 的概率密度为

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

X 的似然函数:

$$L(\theta) = \begin{cases} \frac{1}{\theta^n}, & 0 < x_1, x_2, \dots, x_n \leq \theta, \\ 0, & \text{其他} \end{cases}$$

记 $x_{(n)} = \max(x_1, x_2, \dots, x_n)$, 由于 $x_1, x_2, \dots, x_n \leq \theta$ 等价于 $x_{(n)} \leq \theta$



点估计

用求导方法无法最终确定只能用最大似然原则来求

$$L(\theta) = \begin{cases} \frac{1}{\theta^n}, & \theta \geq x_{(n)}, \\ 0, & \text{其他.} \end{cases}$$

即知当 $\theta < x_{(n)}$ 时 $L(\theta) = 0$;

而当 $\theta \geq x_{(n)}$ 时 $L(\theta)$ 随 θ 的增大而减少.

故 $L(\theta)$ 在 $x_{(n)}$ 取得最大值, 即得 θ 的最大似然估计值为

$$\hat{\theta} = x_{(n)} = \max(x_1, x_2, \dots, x_n).$$



点估计

矩估计法

1. 设 X 是随机变量, 若 $E(X^k), k = 1, 2, \dots$ 存在, 称它为 X 的 \mathbf{k} 阶原点矩, 简称 \mathbf{k} 阶矩.

2. k 阶样本 (原点) 矩

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



点估计

矩估计法是英国统计学家K. 皮尔逊最早提出来的. 由辛钦定理
若总体 X 的数学期望 $E(X) = \mu$ 有限, 则有

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} E(X) = \mu$$

$$A_k = \frac{1}{n} \sum_{i=1}^n X_i^k \xrightarrow{P} \mu_k, k = 1, 2, \dots$$

由依概率收敛的序列的性质知

$$g(A_1, A_2, \dots, A_k) \xrightarrow{P} g(\mu_1, \mu_2, \dots, \mu_k)$$

其中 g 为连续函数



点估计

这表明，当样本容量很大时，在统计上，可以用样本矩去估计总体矩。这一事实导出矩估计法。

定义：用样本原点矩估计相应的总体原点矩，又用样本原点矩的连续函数估计相应的总体原点矩的连续函数，这种参数点估计法称为**矩估计法**。

理论依据：大数定律



点估计

矩估计法的具体做法:

$$\text{令 } \mu_l = A_l, \quad l = 1, 2, \dots, k.$$

这是一个包含 k 个未知参数 $\theta_1, \theta_2, \dots, \theta_k$ 的方程组,

解出其中 $\theta_1, \theta_2, \dots, \theta_k$.

用方程组的解 $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$ 分别作为 $\theta_1, \theta_2, \dots, \theta_k$ 的估计量,
这个估计量称为矩估计量.

矩估计量的观察值称为矩估计值.



点估计

例： 设总体 X 的概率密度为

$$f(x; \mu, \theta) = \begin{cases} \frac{1}{\theta} e^{-(x-\mu)/\theta}, & x > \mu, \\ 0, & \text{其他,} \end{cases}$$

其中 $\mu, \theta (\theta > 0)$ 为待估参数，设 X_1, X_2, \dots, X_n 是来自 X 的一个样本，求 μ, θ 的矩估计量.

解： 总体 X 的一阶、二阶矩分别为

$$\begin{aligned} \mu_1 &= E(X) = \int_{\mu}^{\infty} x \frac{1}{\theta} e^{-(x-\mu)/\theta} dx = \mu + \theta, \\ \mu_2 &= E(X^2) = \int_{\mu}^{\infty} x^2 \frac{1}{\theta} e^{-(x-\mu)/\theta} dx = \mu^2 + 2\theta(\mu + \theta). \end{aligned}$$



点估计

分别以一阶、二阶样本矩 A_1, A_2 代替上两式中的 μ_1, μ_2 , 有

$$\begin{cases} A_1 = \mu + \theta, \\ A_2 = \mu^2 + 2\theta(\mu + \theta). \end{cases}$$

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

从中解得 θ, μ , 即得到 θ, μ 的矩估计量为

$$\hat{\theta} = \sqrt{\frac{1}{n} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right)} = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$$

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



点估计

例： 设总体 X 在 $[0, \theta]$ 上服从均匀分布, 其中 $\theta (\theta > 0)$ 未知, (X_1, X_2, \dots, X_n) 是来自总体 X 的样本, 求 θ 的估计量.

解： 因为 $\mu_1 = E(X) = \frac{\theta}{2},$

根据矩估计法, 令 $\frac{\hat{\theta}}{2} = A_1 = \bar{X},$

所以 $\hat{\theta} = 2\bar{X}$ 为所求 θ 的估计量.



点估计

例： 设总体 X 的均值 μ 和方差 σ^2 都存在, 且有 $\sigma^2 > 0$, 但 μ 和 σ^2 均为未知, 又设 X_1, X_2, \dots, X_n 是一个样本, 求 μ 和 σ^2 的矩估计量.

解： $\mu_1 = E(X) = \mu,$

$$\mu_2 = E(X^2) = D(X) + [E(X)]^2 = \sigma^2 + \mu^2,$$

$$\text{令 } \begin{cases} \mu = A_1, \\ \sigma^2 + \mu^2 = A_2. \end{cases}$$

解方程组得到矩估计量分别为 $\hat{\mu} = A_1 = \bar{X},$

$$\hat{\sigma}^2 = A_2 - A_1^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$



点估计

矩估计法的优缺点

矩估计法的优点是简单易行, 并不需要事先知道总体是什么分布.

缺点是, 当总体类型已知时, 没有充分利用分布提供的信息 .
一般场合下, 矩估计量不具有唯一性 .

其主要原因在于建立矩法方程时, 选取那些总体矩用相应样本矩代替带有一定的随意性 .





3. 估计量的评选标准



估计量的评选标准

问题的提出

对于同一个参数，用不同的估计方法求出的估计量可能不相同。

问题

- (1) 对于同一个参数究竟采用哪一个估计量好？
- (2) 评价估计量的标准是什么？

下面介绍几个常用标准.



估计量的评选标准

在介绍估计量的评选标准之前，我们必须强调指出：

评价一个估计量的好坏，不能仅仅依据一次试验的结果，而必须由多次试验结果来衡量。

这是因为估计量是样本的函数，是随机变量。因此，由不同的观测结果，就会求得不同的参数估计值。因此一个好的估计，应在多次试验中体现出优良性。



估计量的评选标准

常用的几条标准是：

1. 无偏性
2. 有效性
3. 相合性



估计量的评选标准

无偏性

估计量是随机变量，对于不同的样本值会得到不同的估计值。我们希望估计值在未知参数真值附近摆动，而它的期望值等于未知参数的真值。这就导致无偏性这个标准。

设 $\hat{\theta}(X_1, \dots, X_n)$ 是未知参数 θ 的估计量，若

$$E(\hat{\theta}) = \theta$$

则称 $\hat{\theta}$ 为 θ 的无偏估计。



估计量的评选标准

无偏性是对估计量的一个常见而重要的要求。

无偏性的实际意义是指没有系统性的偏差。

例如，用样本均值作为总体均值的估计时，虽无法说明一次估计所产生的偏差，但这种偏差随机地在0的周围波动，对同一统计问题大量重复使用不会产生系统偏差。



估计量的评选标准

例： 设总体 X 的 k 阶矩， $\mu_k = E(X^k)$ 存在. (X_1, X_2, \dots, X_n) 是总体 X 的样本, 试证明不论 X 服从什么分布, $A_k = \frac{1}{n} \sum_{i=1}^n X_i^k$ 是 μ_k 的无偏估计量.

证： 因为 X_1, X_2, \dots, X_n 与 X 同分布,

故有 $E(X_i^k) = E(X^k) = \mu_k, \quad i = 1, 2, \dots, n.$

$$E(A_k) = E\left(\frac{1}{n} \sum_{i=1}^n X_i^k\right) = \frac{1}{n} \sum_{i=1}^n E(X_i^k) = \frac{1}{n} \cdot n \cdot \mu_k = \mu_k$$



估计量的评选标准

例：对于均值 μ , 方差 $\sigma^2 > 0$ 的总体, 若 μ, σ^2 均为未知, 则 σ^2 的估计量 $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ 是有偏的 (不是无偏估计).

解： $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 = A_2 - \bar{X}^2,$

因为 $E(A_2) = \mu_2 = D(X) + [E(X)]^2 = \sigma^2 + \mu^2,$

又因为 $E(\bar{X}^2) = D(\bar{X}) + [E(\bar{X})]^2 = \frac{\sigma^2}{n} + \mu^2,$

所以 $E(\hat{\sigma}^2) = E(A_2 - \bar{X}^2) = E(A_2) - E(\bar{X}^2) = \frac{n-1}{n} \sigma^2 \neq \sigma^2,$

所以 $\hat{\sigma}^2$ 是有偏的.



估计量的评选标准

若以 $\frac{n}{n-1}$ 乘 $\hat{\sigma}^2$, 所得到的估计量就是无偏的.

(这种方法称为**无偏化**).

$$E\left(\frac{n}{n-1}\hat{\sigma}^2\right) = \frac{n}{n-1}E(\hat{\sigma}^2) = \sigma^2.$$

因为

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

即 S^2 是 σ^2 的无偏估计,

故通常取 S^2 作 σ^2 的估计量.



估计量的评选标准

例： 设总体 $X \sim N(\mu, \sigma^2)$, (X_1, X_2, \dots, X_n) 为 X 的一个样本. 求常数 k , 使得 $k \sum_{i=1}^n |X_i - \bar{X}|$ 为 σ 的无偏估计量

解： $E(k \sum_{i=1}^n |X_i - \bar{X}|) = k(\sum_{i=1}^n E|X_i - \bar{X}|)$

注意到 $X_i - \bar{X}$ 是 X_1, X_2, \dots, X_n 的线性函数

$$X_i - \bar{X} = \frac{1}{n}(-X_1 - X_2 \cdots + (n-1)X_i - \cdots - X_n)$$

$$E(X_i - \bar{X}) = 0,$$

$$D(X_i - \bar{X}) = \frac{n+1}{n} \sigma^2$$



估计量的评选标准

$$X_i - \bar{X} \sim N\left(0, \frac{n+1}{n} \sigma^2\right)$$

$$\begin{aligned} E(|X_i - \bar{X}|) &= \int_{-\infty}^{+\infty} |z| \frac{1}{\sqrt{2\pi} \sqrt{\frac{n+1}{n}} \sigma} e^{-\frac{z^2}{2\frac{n+1}{n}\sigma^2}} dz \\ &= 2 \int_0^{+\infty} z \frac{1}{\sqrt{2\pi} \sqrt{\frac{n+1}{n}} \sigma} e^{-\frac{z^2}{2\frac{n+1}{n}\sigma^2}} dz \\ &= \frac{2}{\sqrt{2\pi}} \sqrt{\frac{n+1}{n}} \sigma \end{aligned}$$



估计量的评选标准

故

$$E \left(k \sum_{i=1}^n |X_i - \bar{X}| \right) = k \left(\sum_{i=1}^n E |X_i - \bar{X}| \right) = kn \frac{2}{\sqrt{2\pi}} \sqrt{\frac{n+1}{n}} \sigma$$

$\stackrel{\text{令}}{=} \sigma$

$$k = \sqrt{\frac{\pi}{2n(n+1)}}$$



估计量的评选标准

例： 设总体 X 服从参数为 θ 的指数分布， 概率密度

$$f(x; \theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}}, & x > 0, \\ 0, & \text{其他.} \end{cases}$$

其中参数 $\theta > 0$, 又设 X_1, X_2, \dots, X_n 是来自总体 X 的样本， 试证 \bar{X} 和 $nZ = n[\min(X_1, X_2, \dots, X_n)]$ 都是 θ 的无偏估计.

证： $E(X) = \theta, E(\bar{X}) = \theta$,

所以 \bar{X} 是参数 θ 的无偏估计量.



估计量的评选标准

而 $Z = \min(X_1, \dots, X_n)$ 具有概率密度

$$f_{\min}(x; \theta) = \begin{cases} \frac{n}{\theta} e^{-\frac{nx}{\theta}}, & x > 0, \\ 0, & \text{其它,} \end{cases}$$

故知 $E(Z) = \frac{\theta}{n}$, $E(nZ) = \theta$

即 nZ 也是参数 θ 的无偏估计量 .



估计量的评选标准

由以上例子可知, 一个参数可以有不同的无偏估计量.

比较参数 θ 的两个无偏估计量 $\hat{\theta}_1$ 和 $\hat{\theta}_2$, 如果在样本容量 n 相同的情况下, $\hat{\theta}_1$ 的观察值在真值 θ 的附近较 $\hat{\theta}_2$ 更密集, 则认为 $\hat{\theta}_1$ 较 $\hat{\theta}_2$ 好.

由于方差是随机变量取值与其数学期望的偏离程度, 所以无偏估计以方差小者为好.

这就引进了**有效性**这一概念 .



估计量的评选标准

有效性

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 θ 的无偏估计量, 若有 $D(\hat{\theta}_1) \leq D(\hat{\theta}_2)$,

则称 $\hat{\theta}_1$ 较 $\hat{\theta}_2$ 有效.

例: 设总体 X 服从参数为 θ 的指数分布, 概率密度

$$f(x; \theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}}, & x > 0, \\ 0, & \text{其他.} \end{cases}$$

其中参数 $\theta > 0$, 又设 X_1, X_2, \dots, X_n 是来自总体 X 的样本, \bar{X} 和 $nZ = n[\min(X_1, X_2, \dots, X_n)]$ 都是 θ 的无偏估计. 试证当 $n > 1$ 时, θ 的无偏估计量 \bar{X} 较 nZ 有效.



估计量的评选标准

证： $D(X) = \theta^2$,

故有

$$D(\bar{X}) = D\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n D(X_i) = \frac{\theta^2}{n}$$

而 $D(Z) = \frac{\theta^2}{n^2}$, 故有 $D(nZ) = \theta^2$.

当 $n > 1$ 时, $D(nZ) > D(\bar{X})$,

故 \bar{X} 较 nZ 有效



估计量的评选标准

相合性

若 $\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ 为参数 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 , 则称 $\hat{\theta}$ 为 θ 的相合估计量.

$\hat{\theta}$ 为 θ 的相合估计量



对于任意 $\varepsilon > 0$, 有

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



估计量的评选标准

相合性是对估计量的一个基本要求，不具备相合性的估计量是不予以考虑的。

由辛钦定理

若总体 X 的数学期望 $E(X) = \mu$ 存在，则有

$$A_k = \frac{1}{n} \sum_{i=1}^n X_i^k \xrightarrow{P} E(X^k) = \mu_k \quad (k = 1, 2, \dots)$$

故

$$A_k = \frac{1}{n} \sum_{i=1}^n X_i^k$$

为 $E(X^k) = \mu_k \quad (k = 1, 2, \dots)$ 的相合估计量。



估计量的评选标准

关于相合性的两个常用结论

1. 样本 k 阶矩是总体 k 阶矩的相合性估计量.

由大数定律证明

2. 设 $\hat{\theta}$ 是 θ 的无偏估计量, 且 $\lim_{n \rightarrow \infty} D(\hat{\theta}) = 0$ 则 $\hat{\theta}$ 为 θ 的相合估计量

用切比雪夫不等式证明

矩估计法得到的估计量一般为相合估计量
在一定条件下, 极大似然估计具有相合性



估计量的评选标准

例： $X \sim f(x; \theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & x > 0, \\ 0 & x \leq 0 \end{cases}$, $\theta > 0$ 为常数, 则 \bar{X} 是 θ

的无偏、有效、相合估计量.

证：已证明 \bar{X} 是 θ 的无偏、有效估计量

$$\lim_{n \rightarrow \infty} D(\bar{X}) = \lim_{n \rightarrow \infty} \frac{\theta^2}{n} = 0$$

所以 \bar{X} 是 θ 的相合估计量





4. 区间估计



区间估计

引言

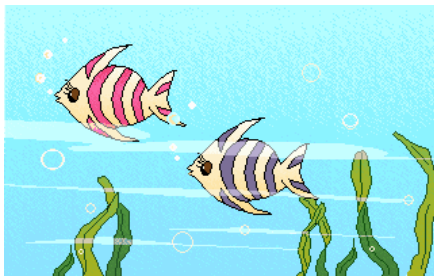
前面，我们讨论了参数点估计. 它是用样本算得的一个值去估计未知参数. 但是，点估计值仅仅是未知参数的一个近似值，它没有反映出这个近似值的误差范围，使用起来把握不大. 区间估计正好弥补了点估计的这个缺陷 .



区间估计

譬如，在估计湖中鱼数的问题中，若我们根据一个实际样本，得到鱼数 N 的极大似然估计为1000条.

实际上， N 的真值可能大于1000条，也可能小于1000条.



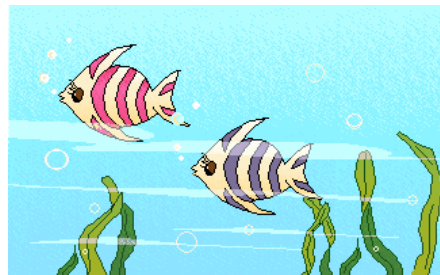
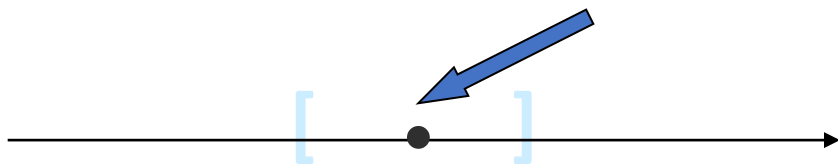
若我们能给出一个区间，在此区间内我们合理地相信 N 的真值位于其中. 这样对鱼数的估计就有把握多了.



区间估计

也就是说，我们希望确定一个区间，使我们能以比较高的可靠程度相信它包含真参数值.

湖中鱼数的真值



这里所说的“可靠程度”是用概率来度量的，
称为置信度或置信水平.

习惯上把置信水平记作 $1 - \alpha$ ，这里 α 是一个很小的正数.



区间估计

置信水平的大小是根据实际需要选定的. 例如, 通常可取置信水平 $1 - \alpha = 0.95$ 或 0.9 等. 根据一个实际样本, 由给定的置信水平, 我们求出一个尽可能小的区间 $(\underline{\theta}, \bar{\theta})$, 使

$$P\{\underline{\theta} < \theta < \bar{\theta}\} \geq 1 - \alpha$$

称区间 $(\underline{\theta}, \bar{\theta})$ 为 θ 的置信水平为 $1 - \alpha$ 置信区间.

对参数 θ 作区间估计, 就是要设法找出两个只依赖于样本的界限 (构造统计量).

$$\underline{\theta} = \underline{\theta}(X_1, X_2, \dots, X_n), \bar{\theta} = \bar{\theta}(X_1, X_2, \dots, X_n), (\underline{\theta} < \bar{\theta})$$

一旦有了样本, 就把 θ 估计在区间 $(\underline{\theta}, \bar{\theta})$ 内.



区间估计

定义： 设总体 X 的分布函数 $F(x; \theta)$ 含有一个未知参数 θ , $\theta \in \Theta$, 对于给定值 α ($0 < \alpha < 1$), 若由来自 X 的样本 X_1, X_2, \dots, X_n 确定的两个统计量 $\underline{\theta} = \underline{\theta}(X_1, X_2, \dots, X_n)$ 和 $\bar{\theta} = \bar{\theta}(X_1, X_2, \dots, X_n)$ ($\underline{\theta} < \bar{\theta}$) 对于任意 $\theta \in \Theta$ 满足

$$P\{\underline{\theta} < \theta < \bar{\theta}\} \geq 1 - \alpha$$

则称随机区间 $(\underline{\theta}, \bar{\theta})$ 是 θ 的置信水平为 $1 - \alpha$ 的**置信区间**, $\underline{\theta}$ 和 $\bar{\theta}$ 称为置信水平为 $1 - \alpha$ 的双侧置信区间的**置信下限**和**置信上限**, $1 - \alpha$ 称为**置信水平**。



区间估计

这里有两个要求：

1. 要求 θ 以很大的可能被包含在区间 $(\underline{\theta}, \bar{\theta})$ 内，就是说，概率 $P\{\underline{\theta} < \theta < \bar{\theta}\}$ 要尽可能大. 即要求估计尽量可靠.
2. 估计的精度要尽可能的高. 如要求区间长度 $\bar{\theta} - \underline{\theta}$ 尽可能短，或能体现该要求的其它准则.

可靠度与精度是一对矛盾，一般是在保证可靠度的条件下尽可能提高精度.



区间估计

置信区间的求法

在求置信区间时，要查表求分位点.

设 $0 < \alpha < 1$, 对随机变量 X , 称满足

$$P(X > x_\alpha) = \alpha \Leftrightarrow P(X \leq x_\alpha) = 1 - \alpha$$

的点 x_α 为 X 的概率分布的上 α 分位点. 继续推导可得

$$P(a < X < b) = 1 - \alpha$$

$$P(X < b) - P(X < a) = 1 - \alpha$$

$$P(X < b) = 1 - \frac{\alpha}{2}, P(X < a) = \frac{\alpha}{2}$$

$$b = x_{\frac{\alpha}{2}}. a = x_{1-\frac{\alpha}{2}},$$



区间估计

例： 设 X_1, \dots, X_n 是取自 $N(\mu, \sigma^2)$ 的样本， σ^2 已知，求参数 μ 的置信度为 $1 - \alpha$ 的置信区间。

解： 选 μ 的点估计为 \bar{X} ，

寻找未知参数的一个良好估计。

取

$$U = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

寻找一个待估参数和统计量的函数，要求其分布为已知。

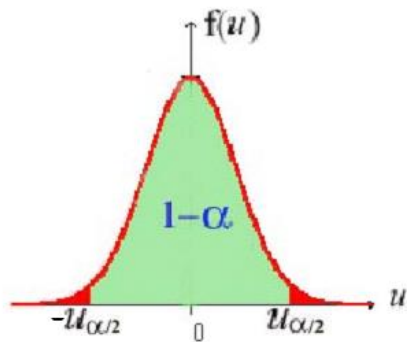


区间估计

对给定的置信水平 $1 - \alpha$

查正态分布表得 $u_{\alpha/2}$, 使得

$$P\left\{\left|\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right| \leq u_{\alpha/2}\right\} = 1 - \alpha$$



$$P\left\{\bar{X} - \frac{\sigma}{\sqrt{n}}u_{\alpha/2} \leq \mu \leq \bar{X} + \frac{\sigma}{\sqrt{n}}u_{\alpha/2}\right\} = 1 - \alpha$$

于是所求 μ 的置信区间为

$$\left[\bar{X} - \frac{\sigma}{\sqrt{n}}u_{\alpha/2}, \bar{X} + \frac{\sigma}{\sqrt{n}}u_{\alpha/2}\right]$$

简记为 $(\bar{X} \pm \frac{\sigma}{\sqrt{n}}u_{\frac{\alpha}{2}})$

区间估计

区间估计的5个步骤

(1) 取 θ 的较优点估计量 $\hat{\theta}(X_1, X_2, \dots, X_n)$

(2) 由 $\hat{\theta}$ 出发, 找一个样本函数 $W = W(\hat{\theta}, \theta)$ 其分布已知, 且只含一个未知参数 θ

(3) 查表求得 W 的分位点, 使得

$$P(W_{\frac{\alpha}{2}} < W(\theta) < W_{1-\frac{\alpha}{2}}) = 1 - \alpha$$

(4) 反解 (3) 中的不等式, 即得 $P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1 - \alpha$

(5) 将样本值代入 $\hat{\theta}_1, \hat{\theta}_2$, 得置信区间 $(\hat{\theta}_1, \hat{\theta}_2)$



区间估计

需要指出的是, 给定样本, 给定置信水平, 置信区间也不是唯一的. 对同一个参数, 我们可以构造许多置信区间. 以上一题为例.

例: 设 X_1, \dots, X_n 是取自 $N(\mu, \sigma^2)$ 的样本, σ^2 已知, 求参数 μ 的置信度为 $1 - \alpha$ 的置信区间.

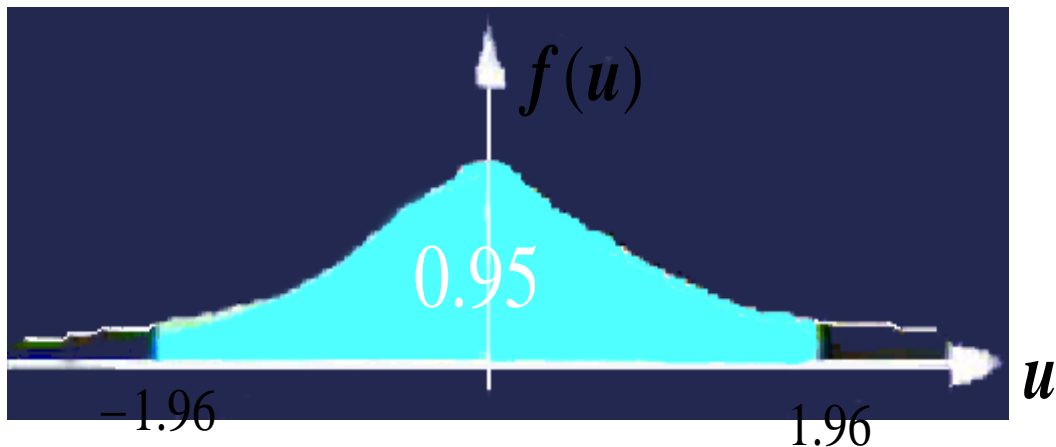
$$U = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

由标准正态分布表, 对任意 a 、 b , 我们可以求得
 $P(a < U < b)$.



区间估计

由 $P(-1.96 \leq U \leq 1.96) = 0.95$

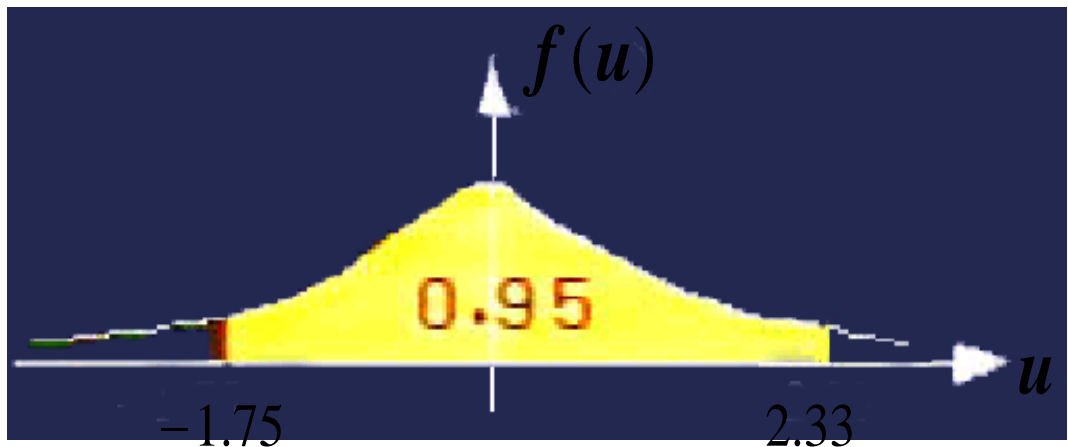


我们得到均值 μ 的置信水平为 $1 - \alpha$ 置信区间为

$$[\bar{X} - 1.96\sigma/\sqrt{n}, \bar{X} + 1.96\sigma/\sqrt{n}]$$

区间估计

由 $P(-1.75 \leq U \leq 2.33) = 0.95$



我们得到均值 μ 的置信水平为 $1 - \alpha$ 置信区间为

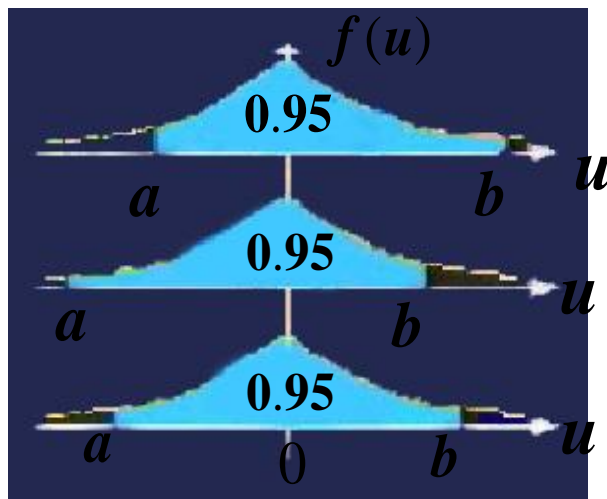
$$[\bar{X} - 1.75\sigma/\sqrt{n}, \bar{X} + 2.33\sigma/\sqrt{n}]$$

这个区间比前面一个要长一些.

区间估计

类似地，我们可得到若干个不同的置信区间。

任意两个数 a 和 b ，只要它们的纵标包含 $f(u)$ 下95%的面积，就确定一个95%的置信区间。

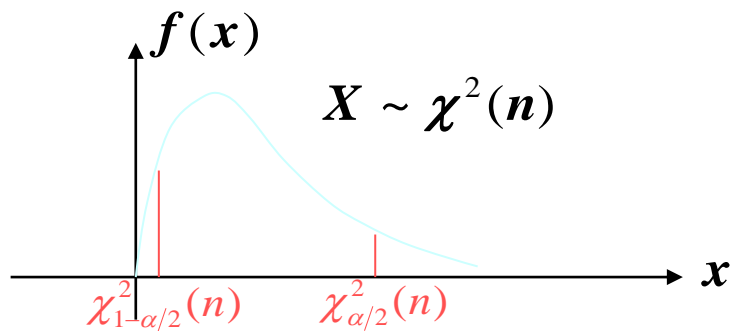


我们总是希望置信区间尽可能短。

区间估计

在概率密度为单峰且对称的情形，当 $a = -b$ 时求得的置信区间的长度为最短。

即使在概率密度不对称的情形，如 χ^2 分布， F 分布，习惯上仍取对称的分位点来计算未知参数的置信区间。





5. 正态总体均值与方差的 区间估计



正态总体均值与方差的区间估计

单个正态总体参数的区间估计

设总体 $X \sim N(\mu, \sigma^2)$, X_1, X_2, \dots, X_n 是 X 的一个样本。

- 1. μ 的区间估计
 - (1) σ^2 已知
 - (2) σ^2 未知
- 2. σ^2 的区间估计
 - (1) μ 已知
 - (2) μ 未知



正态总体均值与方差的区间估计

1. μ 的区间估计

(1) σ^2 已知

采用枢轴量 $u = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ 已经得到 μ 的置信度为 $1 - \alpha$ 的
置信区间为

$$(\bar{X} \pm \frac{\sigma}{\sqrt{n}} u_{\frac{\alpha}{2}})$$

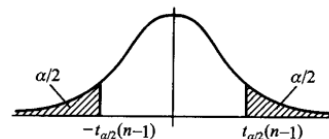


正态总体均值与方差的区间估计

1. μ 的区间估计(2) σ^2 未知

考虑到 S^2 是 σ^2 的无偏估计, 由定理知, $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

并且该分布 $t(n-1)$ 不依赖于任何未知参数



$$P\left(-t_{\frac{\alpha}{2}}(n-1) < \frac{\bar{X}-\mu}{S/\sqrt{n}} < t_{\frac{\alpha}{2}}(n-1)\right) = 1 - \alpha$$

$$\text{即 } P\left(\bar{X} - \frac{S}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1) < \mu < \bar{X} + \frac{S}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1)\right) = 1 - \alpha$$

故 μ 的置信度为 $1 - \alpha$ 的置信区间

$$\left(\bar{X} \pm \frac{S}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1)\right)$$



正态总体均值与方差的区间估计

例：有一大批糖果，现随机地从中取16袋，称得重量（单位：g）如下

506, 508, 499, 503, 504, 510, 497, 512
514, 505, 493, 496, 506, 502, 509, 496

设袋装糖果的重量近似地服从正态分布，试求总体均值 μ 的置信水平为0.95的置信区间.

解：这里 $1 - \alpha = 0.95$ ，查表得 $t_{0.025}(15) = 2.1315$,

由给出的数据公式算得： $\bar{X} = 503.75, S = 6.2022$.



正态总体均值与方差的区间估计

μ 的置信水平为0.95的置信区间为

$$\left(\bar{X} \pm \frac{S}{\sqrt{n}} t_{\frac{\alpha}{2}}(n-1) \right)$$

$$\left(503.75 \pm \frac{6.2022}{\sqrt{16}} \times 2.1315 \right)$$

即 (500.4, 507.1)，这就是说估计袋装糖果重量的均值在500.4g与507.1g之间，这个估计的可信程度为95%。

在实际问题中，总体方差未知的情况居多。

