第 7 章 参数估计

安徽财经大学

统计与应用数学学院

目录

章前导读

- 上一章已经提到, 数理统计的基本问题就是根据样本提供的信息来推断总体的分布或分布的数字特征, 即统计推断.
- 统计推断的主要内容包括参数估计和统计假设检验. 它们是数理统计的核心内容.
- 本章首先讨论了参数的点估计、区间估计及其估计量的评判标准, 然后简要介绍了 Bayes 估计的基本思想和确定先验分布的一些方法.

章前导读

参数估计 (parameter estimation) 包含点估计 (point estimation) 和区间估计 (interval estimation).
所谓参数的点估计就是根据来自总体 X 的样本 $(X_1, X_2, \cdots, X_n)$, 通过构造一个合适的统计量

$$\hat{\theta}\left(X_1, X_2, \cdots, X_n\right),$$

对总体 X 的某些特征或总体分布 $X \sim F(x, \theta)$ 中的未知参数 θ 进行估计, 再把相应的样本观测值 $(x_1, x_2, \cdots, x_n)$ 代入 $\hat{\theta}\left(x_1, x_2, \cdots, x_n\right)$ 作为 θ 的估计值.
因此要讨论参数 θ 的点估计问题关键是如何选取估计量 $\hat{\theta}\left(X_1, X_2, \cdots, X_n\right)$. 矩估计法和极大似然估计法就是两种常用的构造估计量的方法.

1. 点估计
   - 矩估计
   - 极大似然估计法
   - 估计量评判标准
2. 区间估计

矩估计法

- 矩估计法 (moment method of estimation) 是由英国统计学家皮尔逊 (K. Pearson) 于 1894 年提出的一种古老的估计方法.
- 矩估计的基本思想是用样本的经验分布和样本矩作为总体相应矩的估计, 以样本矩的函数作为总体相应矩的同一函数的估计, 其理论依据就是我们前面所学过的大数定律和格里汶科定理.

设总体 $X \sim F(x, \theta_1, \cdots, \theta_k)$, 其中 $(\theta_1, \cdots, \theta_k)$, $k \geqslant 1$ 为 k 个未知参数, $(\theta_1, \cdots, \theta_k)$ 的一切取值范围称为参数空间, 记为 Θ. $(X_1, X_2, \cdots, X_n)$ 是来自总体 X 的一个样本, $(x_1, x_2, \cdots, x_n)$ 为一组样本观测值, 假定总体 X 的 k 阶矩存在, 则其 $r(r \leqslant k)$ 阶矩也存在. 即

$$m_r(\theta_1, \cdots, \theta_k) = EX^r = \int_{-\infty}^{+\infty} x^r \mathrm{d}F(x, \theta_1, \cdots, \theta_k), \quad r = 1, 2, \cdots, k$$

记 r 阶样本矩为

$$A_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r, \quad r = 1, 2, \cdots, k.$$

得到 k 个方程

$$m_r(\theta_1, \cdots, \theta_k) = A_r, r = 1, 2, \cdots, k. \tag{7.1.1}$$

求解方程组, 得到 $\theta_1, \cdots, \theta_k$ 的一组解

$$\hat{\theta}_j = \hat{\theta}_j(X_1, X_2, \cdots, X_n), \quad j = 1, 2, \cdots, k.$$

称 $\hat{\theta}_j(X_1, X_2, \cdots, X_n)$ 为参数 $\theta_j (j = 1, 2, \cdots, k)$ 的**矩估计量**, $\hat{\theta}_j(x_1, x_2, \cdots, x_n)$ 称为参数 θ_j 的**矩估计值**. 这种求参数 θ 的估计量 (值) 的方法称为矩估计法.

例 (7.1.1)

设总体 X 的期望 $E(X)=\mu$, 方差 $D(X)=\sigma^2$, 求 μ 和 σ^2 的矩估计量.

例 (7.1.1)

设总体 X 的期望 $E(X)=\mu$, 方差 $D(X)=\sigma^2$, 求 μ 和 σ^2 的矩估计量.

解

设 $(X_1, X_2, \cdots, X_n)$ 为来自总体 X 的样本, 于是

$$\begin{cases} m_1 = E(X) = \mu, \\ m_2 = E\left(X^2\right) = \mu^2 + \sigma^2, \end{cases}$$

根据 (7.1.1) 式可得 $\begin{cases} \mu = \overline{X}, \\ \mu^2 + \sigma^2 = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} X_i^2, \end{cases}$

解得 $\begin{cases} \hat{\mu} = \overline{X}, \\ \hat{\sigma}^2 = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 = S_n^2, \end{cases}$

即 μ 和 σ^2 的矩估计量分别为样本均值与样本方差.

例 (7.1.2)

设总体 $X \sim U(a, b)$, 其中 $-\infty < a < b < +\infty$, 求 a, b 的矩估计量.

例 (7.1.2)

设总体 $X \sim U(a, b)$, 其中 $-\infty < a < b < +\infty$, 求 a, b 的矩估计量.

解

因为 $m_1 = \int_a^b \frac{x}{b-a}\,\mathrm{d}x = \frac{a+b}{2}$, $m_2 = \int_a^b \frac{x^2}{b-a}\,\mathrm{d}x = \frac{a^2+ab+b^2}{3}$, 根据 (7.1.1) 式可得

$$\begin{cases} \dfrac{a+b}{2} = \overline{X}, \\ \dfrac{a^2+ab+b^2}{3} = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i^2. \end{cases}$$

解方程组可得 a, b 的矩估计量分别为 $\hat{a} = \overline{X} - \sqrt{3}S_n$, $\hat{b} = \overline{X} + \sqrt{3}S_n$.

矩估计

- 矩估计法是一种比较简单的估计法，利用矩估计时不需要知道总体的具体分布，只要知道总体的相应阶矩即可. 如例 7.1.1 中，只需知道总体均值和方差，就可求出其相应的矩估计量.
- 矩估计量有时并不唯一，如总体 $X \sim P(\lambda)$，由于泊松分布的均值和方差都为 λ，根据例 7.1.1，可用 $\overline{X}$ 或 S_n^2 作为 λ 的矩估计量.
- 矩估计法对信息利用不够充分，这是估计法的不足之处. 这种情况一般选低阶矩 $\overline{X}$ 作为 λ 的估计量.

极大似然估计

极大似然估计 (maximum likelihood estimation) 是在总体分布概率密度已知的情况下，求参数 θ 估计量的另一种常用的方法. 它最早是由高斯提出来的，后来 Fisher 在 1912 年重新提出并证明了它的一些相关性质，其基本思想是极大似然原理.

例 (7.1.3)

设一个箱子中有很多大小相同的红球和黑球，只知道这两种球数目比为 $3:1$，而不知道到底哪种颜色球多，为估计箱子中到底哪种颜色球较多，现从中有放回抽取 3 个球，试根据抽样结果来判断箱子中哪种颜色球较多?

解

设 X 表示抽取 3 个球中红球的个数，p 为每次抽到红球的概率，则 $X \sim b(3, p)$，即

$$P(X = k) = b(k; 3, p) = C_3^k p^k (1-p)^{3-k}, k = 0, 1, 2, 3.$$

问题归结为对未知参数 p 的估计，即 $p = \dfrac{1}{4}$ 还是 $p = \dfrac{3}{4}$? 根据抽样的结果，其概率有下表中 4 种可能的情形.

X	0	1	2	3
$b(k; 3, 1/4)$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$
$b(k; 3, 3/4)$	$\frac{1}{64}$	$\frac{9}{64}$	$\frac{27}{64}$	$\frac{27}{64}$

解

若 3 次抽球中没有抽到红球，即 $k=0$，当 $p=\frac{1}{4}$ 时，$P(X=0)=\frac{27}{64}$；当 $p=\frac{3}{4}$ 时，$P(X=0)=\frac{1}{64}$. 表明随机事件 $\{X=0\}$ 在 $p=\frac{1}{4}$ 时发生的可能性比 $p=\frac{3}{4}$ 时发生的可能性更大，因此选 $\frac{1}{4}$ 作为 p 的估计值更合理. 其它情形可类似讨论，从而可得参数 p 的合理估计为

$$\hat{p}=\begin{cases} 1/4, & \text{当} k=0,1; \\ 3/4, & \text{当} k=2,3. \end{cases}$$

上述选取 p 的估计值 $\hat{p}$ 的原则是：当 $\hat{p}$ 作为 p 的估计值时，使样本观测值出现的概率最大. 这就是所谓的极大似然原理.

以下分别从离散型和连续型总体两种情形来介绍极大似然估计法.

(1) 设总体 X 为离散型随机变量, 其分布列为 $P(X=x)=p(x,\theta)$, 其中 $\theta\in\Theta$ 为待估的未知参数. $(x_1,x_2,\cdots,x_n)$ 为样本 $(X_1,X_2,\cdots,X_n)$ 的一组观测值, 则其出现的概率为

$$P(X_1=x_1,X_2=x_2,\cdots,X_n=x_n)=\prod_{i=1}^{n}P(X_i=x_i)=\prod_{i=1}^{n}p(x_i,\theta),$$

$\prod\limits_{i=1}^{n}p(x_i,\theta)$ 为 θ 的函数, 记为 $L(\theta)=\prod\limits_{i=1}^{n}p(x_i,\theta)$, 称为该样本的似然函数.

极大似然原理就是选取 θ 的估计值 $\hat{\theta}=\hat{\theta}(x_1,x_2,\cdots,x_n)$ 使得似然函数 $L(\theta)$ 达到最大值.

$$\prod_{i=1}^{n}p(x_i,\hat{\theta})=\sup_{\theta\in\Theta}L(\theta)=\sup_{\theta\in\Theta}\prod_{i=1}^{n}p(x_i,\theta). \tag{7.1.2}$$

(2) 设总体 X 为连续型随机变量, 其概率密度为 $f(x,\theta), \theta \in \Theta$ 为待估的未知参数. 则样本 $(X_1, X_2, \cdots, X_n)$ 的联合密度为
$\prod_{i=1}^{n} f(x_i, \theta)$, $(x_1, x_2, \cdots, x_n)$ 为一组样本观测值, 于是样本落入 $(x_1, x_2, \cdots, x_n)$ 的邻域内的概率为 $\prod_{i=1}^{n} f(x_i, \theta)\,\Delta x_i$, 也为 θ 的函数. 根据极大似然原理, 选取 θ 的估计值 $\hat{\theta} = \hat{\theta}(x_1, x_2, \cdots, x_n)$ 使得

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

由于 Δx_i 是不依赖于 θ 的增量, 故我们只需求出 $\hat{\theta} = \hat{\theta}(x_1, x_2, \cdots, x_n)$ 使得

$$\prod_{i=1}^{n} f(x_i, \hat{\theta}) = \sup_{\theta\in\Theta} \prod_{i=1}^{n} f(x_i, \theta)\,. \tag{7.1.3}$$

同样称 $L(\theta) = \prod_{i=1}^{n} f(x_i, \theta)$ 为似然函数.

由此可见，不管总体是离散型还是连续型，我们只需知道其分布列或概率密度函数，就可以得到一个似然函数 $L(\theta)$，其中 θ 为任意 k 维的参数向量 $\theta=(\theta_1\cdots,\theta_k)$；然后根据极大似然原理来寻找 θ 的估计值 $\hat{\theta}=\hat{\theta}\left(x_1,x_2,\cdots,x_n\right)$ 使得

$$L\left(\hat{\theta};x_1,\cdots,x_n\right)=\sup_{\theta\in\Theta}L\left(\theta;x_1,\cdots,x_n\right). \tag{7.1.4}$$

如果存在 $\hat{\theta}\left(X_1,X_2,\cdots,X_n\right)$ 使得 (7.1.4) 式成立，则称之为 θ 的**极大似然估计量**，记为 $\hat{\theta}_{\mathrm{MLE}}$，相应的估计值 $\hat{\theta}\left(x_1,x_2,\cdots,x_n\right)$ 为 θ 的极大似然估计值. 上述这种求参数 θ 的估计量的方法称为**极大似然估计法**.

由于 $\ln L(\theta)$ 是 $L(\theta)$ 的单增函数，故 $L(\theta)$ 与 $\ln L(\theta)$ 在同一点取得极大值，于是 (7. 1.4) 式可转化为

$$\ln L\left(\hat{\theta}; x_1, \cdots, x_n\right) = \sup_{\theta \in \Theta} \ln L\left(\theta; x_1, \cdots, x_n\right). \tag{7.1.5}$$

为求 (7.1.5) 式的解，根据多元微分学知识，若 Θ 为开集，似然函数 $L(\theta)$ 关于 θ 的偏导数存在，则 $\theta = (\theta_1, \cdots, \theta_k)$ 的极大似然估计 $\hat{\theta} = \left(\hat{\theta}_1, \cdots, \hat{\theta}_k\right)$ 一定为方程组

$$\frac{\partial}{\partial \theta_i} L\left(\theta_1, \cdots, \theta_k\right) = 0, i = 1, 2, \cdots, k \tag{7.1.6}$$

或

$$\frac{\partial}{\partial \theta_i} \ln L\left(\theta_1, \cdots, \theta_k\right) = 0, i = 1, 2, \cdots, k \tag{7.1.7}$$

的解. (7.1.6) 式或 (7.1.7) 式称为似然方程.

例 (7.1.4)

设总体 $X \sim N(\mu, \sigma^2)$, 其中 μ, σ^2 为未知参数, $(X_1, \cdots, X_n)$ 为来自总体 X 的样本, 求 μ, σ^2 的极大似然估计.

例 (7.1.4)

设总体 $X \sim N(\mu, \sigma^2)$, 其中 μ, σ^2 为未知参数, $(X_1, \cdots, X_n)$ 为来自总体 X 的样本, 求 μ, σ^2 的极大似然估计.

解

正态分布的似然函数为

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

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

解

故似然方程为

$$\begin{cases} \dfrac{\partial \ln L\left(\mu, \sigma^2\right)}{\partial \mu} = \dfrac{1}{\sigma^2} \sum\limits_{i=1}^{n} (x_i - \mu) = 0, \\ \dfrac{\partial \ln L\left(\mu, \sigma^2\right)}{\partial \sigma^2} = -\dfrac{n}{2\sigma^2} + \dfrac{1}{2\sigma^4} \sum\limits_{i=1}^{n} (x_i - \mu)^2 = 0, \end{cases}$$

解方程组得 $\begin{cases} \mu = \dfrac{1}{n} \sum\limits_{i=1}^{n} x_i = \overline{x}, \\ \sigma^2 = \dfrac{1}{n} \sum\limits_{i=1}^{n} (x_i - \overline{x})^2 = s_n^2, \end{cases}$

于是 μ, σ^2 的极大似然估计量为

$$\begin{cases} \hat{\mu} = \dfrac{1}{n} \sum\limits_{i=1}^{n} X_i = \overline{X}, \\ \hat{\sigma}^2 = \dfrac{1}{n} \sum\limits_{i=1}^{n} \left(X_i - \overline{X}\right)^2 = S_n^2. \end{cases}$$

例 (7.1.5)

投掷一枚硬币, 假设出现正面朝上的概率 p 未知. 现独立地投掷 100 次, 发现出现 48 次正面朝上, 试求 p 的极大似然估计值.

例 (7.1.5)

投郑一枚硬币，假设出现正面朝上的概率 p 未知. 现独立地投掷 100 次，发现出现 48 次正面朝上，试求 p 的极大似然估计值.

解

设 X 表示一次试验中出现正面的次数，则 X 服从参数为 p 的两点分布. 即 X 的分布列为 $P(X=x)=p^x(1-p)^{1-x}$, $x=0,1$.
令 X_i, $i=1,2,\cdots,100$ 为来自总体 X 的样本，x_i, $i=1,2,\cdots,100$ 为一组样本观测值，则似然函数为

$$L(p)=\prod_{i=1}^{100}p^{x_i}(1-p)^{1-x_i}=p^{\sum\limits_{i=1}^{100}x_i}(1-p)^{100-\sum\limits_{i=1}^{100}x_i}.$$

例 (7.1.5)

投郑一枚硬币，假设出现正面朝上的概率 p 未知．现独立地投掷 100 次，发现出现 48 次正面朝上，试求 p 的极大似然估计值．

解

于是似然方程为

$$\frac{\mathrm{d}\ln L(p)}{\mathrm{d}p}=\frac{\sum\limits_{i=1}^{100}x_i}{p}-\frac{100-\sum\limits_{i=1}^{100}x_i}{1-p}=0,$$

解以上方程得 $p=\sum\limits_{i=1}^{100}x_i/100$．由于 $\sum\limits_{i=1}^{100}x_i=48$，从而 p 的极大似然估计值为 $\hat{p}=0.48$．

例 (7.1.6)

设 $X \sim U[0,\theta], \theta > 0$，其中 θ 为未知参数，$(X_1, \cdots, X_n)$ 为来自总体 X 的样本，求 θ 的矩估计与极大似然估计.

例 (7.1.6)

设 $X \sim U[0,\theta], \theta > 0$，其中 θ 为未知参数，$(X_1, \cdots, X_n)$ 为来自总体 X 的样本，求 θ 的矩估计与极大似然估计.

解

因为 $EX = \theta/2$，根据矩估计法，令 $EX = \theta/2 = \overline{X}$，于是 θ 的矩估计量为

$$\hat{\theta} = 2\overline{X}.$$

由于

$$f(x,\theta) = \begin{cases} 1/\theta, & x \in [0,\theta]; \\ 0, & x \notin [0,\theta]; \end{cases}$$

故似然函数为

$$L(\theta) = \begin{cases} 1/\theta^n, & x_i \in [0,\theta],\ i = 1, 2, \cdots, n; \\ 0, & \text{其他}. \end{cases}$$

例 (7.1.6)

设 $X \sim U[0,\theta], \theta > 0$, 其中 θ 为未知参数, $(X_1, \cdots, X_n)$ 为来自总体 X 的样本, 求 θ 的矩估计与极大似然估计.

解

由于似然函数不是处处可微, 不能用求导数方法求解, 但可以直接利用极大似然原理的思想来求解. 要使似然函数取非零值, 且 $L(\theta)$ 达到最大, 则 θ 必须尽可能小, 而 $\theta \geqslant x_i$, $i = 1, 2, \cdots, n$, 故 θ 的极大似然估计量为 $\hat{\theta} = \max\limits_{1 \leqslant i \leqslant n} \{X_i\}$.

1. 点估计
   - 矩估计
   - 极大似然估计法
   - 估计量评判标准
2. 区间估计

估计量的评判标准

- 从例 7.1.6 可知，同一参数可以存在不同的估计量，到底哪一个好些？
- 另外，由于估计量作为样本的函数是一个随机变量，从而对不同的样本观测值，相应的估计量的观测值也会不同，因此要考察一个估计量的好坏就要进行多次观察，从其统计规律来判断，而不能仅凭一次观测结果来判断.
- 本节从估计量的无偏性、有效性和一致性三个方面来讨论衡量估计量优良性的标准.

1. 无偏性

定义 (7.1.1)

设 $\hat{\theta}=\hat{\theta}(X_1,\cdots,X_n)$ 为未知参数 $\theta\in\Theta$ 的估计量. 如果对一切 $\theta\in\Theta$, 都有

$$E\hat{\theta}=\theta.$$

则称 $\hat{\theta}$ 为参数 θ 的无偏估计量 (unbiased estimator).

估计量 $\hat{\theta}$ 为样本的函数, 它是随机变量, 其取值随样本观测值的改变而改变. (7.1.8) 表明 $\hat{\theta}$ 的取值围绕 θ 波动, 但就其平均来说为 θ 的真值. 无偏性体现了一种频率的思想, 只有在大量重复抽样时, 才有意义. 一般来说, 一个参数的无偏估计可以不唯一, 有时还可能不存在. 由矩估计法和极大似然估计法求出的估计量, 也不一定为无偏估计量.

如例 7.1.1 中用样本均值 $\overline{X}$ 估计总体均值 $EX = \mu$，样本方差 S_n^2 作为总体方差 σ^2 的估计量，根据第 6 章知识，我们有 $E\overline{X} = \mu$，
$ES_n^2 = \frac{n-1}{n}\sigma^2 \neq \sigma^2$，可见样本均值为总体均值的无偏估计，而样本方差不为总体方差的无偏估计.
作适当修正，令

$$S^2 = \frac{n}{n-1} S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2 ,$$

则 S^2 为 σ^2 的无偏估计量.

例 (7.1.7)

判别例 7.1.6 中两种方法求出的估计量是否为参数 θ 的无偏估计?

例 (7.1.7)

判别例 7.1.6 中两种方法求出的估计量是否为参数 θ 的无偏估计?

解

由例 7.1.6 知 θ 的矩估计量为 $\hat{\theta}=2\overline{X}$, 而 $E\hat{\theta}=E(2\overline{X})=2\cdot\dfrac{\theta}{2}=\theta$, 故 $\hat{\theta}=2\overline{X}$ 为 θ 的无偏估计量. θ 的极大似然估计量为 $\hat{\theta}=\max\limits_{1\leqslant i\leqslant n}\{X_i\}$, 根据极大次序统计量分布可得

$$F_{\hat{\theta}}(x)=P\left(\max_{1\leqslant i\leqslant n}X_i\leqslant x\right)=[F_X(x)]^n=\begin{cases}0, & x<0;\\ \dfrac{x^n}{\theta^n}, & 0\leqslant x<\theta;\\ 1, & x\geqslant\theta.\end{cases}$$

其中 $F_X(x)$ 为总体 X 的分布函数.

例 (7.1.7)

判别例 7.1.6 中两种方法求出的估计量是否为参数 θ 的无偏估计?

解

由此可得 $\hat{\theta}$ 的密度函数为

$$f_{\hat{\theta}}(x)=\begin{cases}\dfrac{nx^{n-1}}{\theta^n}; & 0<x\leqslant\theta;\\ 0, & \text{其他}.\end{cases}$$

于是

$$E\hat{\theta}=\int_0^{\theta}xf_{\hat{\theta}}(x)\mathrm{d}x=\int_0^{\theta}x\frac{nx^{n-1}}{\theta^n}\,\mathrm{d}x=\frac{n}{n+1}\theta\neq\theta.$$

从而 $\hat{\theta}=\max\limits_{1\leqslant i\leqslant n}\{X_i\}$ 不为参数 θ 的无偏估计量, 适当修正 $\hat{\theta}^*=\dfrac{n+1}{n}\max\limits_{1\leqslant i\leqslant n}\{X_i\}$ 就为参数 θ 的无偏估计量.

2. 有效性

若一个参数有多个无偏估计，这些无偏估计量中哪个更好呢？因为方差是反映随机变量取值的离散程度，在期望相同时，自然希望其方差尽可能小．下面引入有效性的概念．

定义 (7.1.2)

设 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 都为参数 θ 的无偏估计量，若对任意参数 $\theta \in \Theta$ 和固定的样本容量 n，有

$$D(\hat{\theta}_1) \leqslant D(\hat{\theta}_2),$$

则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 更有效．

例 (7.1.8)

对例 7.1.6 中 θ 的两个无偏估计 $\hat{\theta}_1 = 2\overline{X}$ 和 $\hat{\theta}^* = \dfrac{n+1}{n}\max\limits_{1\leqslant i\leqslant n}\{X_i\}$, 比较哪个更有效?

例 (7.1.8)

对例 7.1.6 中 θ 的两个无偏估计 $\hat{\theta}_1 = 2\overline{X}$ 和 $\hat{\theta}^* = \dfrac{n+1}{n} \max\limits_{1\leqslant i\leqslant n} \{X_i\}$, 比较哪个更有效?

解

$$D\hat{\theta}_1 = 4D\overline{X} = 4 \times \frac{\theta^2}{12n} = \frac{\theta^2}{3n}.$$

$$D\hat{\theta}^* = D\left[\frac{n+1}{n} \max_{1\leqslant i\leqslant n} \{X_i\}\right] = \left(\frac{n+1}{n}\right)^2 E\left[\max_{1\leqslant i\leqslant n} \{X_i\}\right]^2 - \theta^2$$

$$= \left(\frac{n+1}{n}\right)^2 \int_0^\theta x^2 \frac{nx^{n-1}}{\theta^n}\,\mathrm{d}x - \theta^2 = \frac{\theta^2}{n(n+2)},$$

当 $n > 1$ 时, $D\hat{\theta}^* = \dfrac{\theta^2}{n(n+2)} < \dfrac{\theta^2}{3n} = D\hat{\theta}_1$. 因此 $\hat{\theta}^*$ 比 $\hat{\theta}_1$ 更有效.

3. 一致性

无偏性和有效性是在样本容量 n 固定的情形下来评判估计量的性质，而估计量不仅与样本的取值有关，还同样本容量 n 有关．一般地，总体的信息随着样本容量的增加也随之增加，从而估计量取值越来越接近真值．

定义 (7.1.3)

设 $\hat{\theta}_n(X_1, X_2, \cdots, X_n)$ 为未知参数 θ 的估计量，若 $\forall \varepsilon > 0$，有

$$\lim_{n\to\infty} P\left(\left|\hat{\theta}_n - \theta\right| \geqslant \varepsilon\right) = 0 \text{ 或 } \lim_{n\to\infty} P\left(\left|\hat{\theta}_n - \theta\right| < \varepsilon\right) = 1$$

成立，则称 $\hat{\theta}_n$ 为参数 θ 的一致估计量 (uniformly estimator).

根据辛钦大数定律，我们可证明样本均值为总体均值的一致估计量；样本方差为总体方差的一致估计量；样本的 k 阶矩为总体 k 阶矩的一致估计量．

1 点估计

2 区间估计
- 单个正态总体参数的区间估计
- 两个正态总体参数的区间估计

上一节讨论了参数的点估计，即用一个估计量 $\hat{\theta}$ 来估计未知参数 θ，当给定一组样本观测值时，得到 θ 的一个估计值．但由于 $\hat{\theta}$ 为随机变量，它取参数 θ 的真值可能性很小，并且点估计没有给出误差的界限．因此我们希望给出参数 θ 的一个范围，同时使得 θ 落入该范围的概率比较大，这就要引入参数的区间估计．

定义 (7.2.1)

设总体 $X \sim F(x,\theta), \theta$ 为未知参数，$(X_1, X_2, \cdots, X_n)$ 是来自总体 X 的样本，$\hat{\theta}_1(X_1, X_2, \cdots, X_n)$，$\hat{\theta}_2(X_1, X_2, \cdots, X_n)$ 是两个统计量，若对一切样本和给定的概率 $1-\alpha$ $(0<\alpha<1)$ 有

$$P\left\{\hat{\theta}_1(X_1, X_2, \cdots, X_n) < \theta < \hat{\theta}_2(X_1, X_2, \cdots, X_n)\right\} = 1-\alpha.$$

则称随机区间 $(\hat{\theta}_1, \hat{\theta}_2)$ 为参数 θ 的 $1-\alpha$ –**置信区间** (confidence interval)，$\hat{\theta}_1, \hat{\theta}_2$ 分别称为置信下限与置信上限，$1-\alpha$ 称为**置信度或置信水平** (confidence level).

- 在经典统计中, 认为参数 θ 为一个确定值, 因此 (7.2.1) 式的意义是指随机区间 $(\hat{\theta}_1, \hat{\theta}_2)$ 以 $1-\alpha$ 的概率套住 θ. 即在多次重复抽样中所得到的置信区间中, 大约有 $(1-\alpha)\times 100\%$ 的区间包含 θ.
- $1-\alpha$ 表示置信区间的可靠性; 固定 α, 区间长度 $\hat{\theta}_2-\hat{\theta}_1$ 反映了置信区间的精度.
- 一般来说, α 和 $\hat{\theta}_2-\hat{\theta}_1$ 越小, 则相应的估计越好, 但当样本容量 n 固定时, α 和 $\hat{\theta}_2-\hat{\theta}_1$ 不可能同时都很小, 减小其中一个就会使另一个增大.
- 实际中, 我们是在保证可靠性的前提下, 努力提高精度.

设总体 $X \sim N\left(\mu, \sigma^2\right), (X_1, X_2, \cdots, X_n)$ 是来自总体 X 的样本, 下面讨论参数 μ, σ^2 的区间估计.

1. 参数 μ 的区间估计

(1) 设 $\sigma^2=\sigma_0^2$ 已知. 令

$$U=\frac{\overline{X}-\mu}{\sigma_0/\sqrt{n}},$$

由 Fisher 定理知 $U\sim N(0,1)$, 对给定的置信水平 $1-\alpha$, 查正态分布表可得 $P\left\{|U|<u_{\alpha/2}\right\}=1-\alpha$, 即

$$P\left\{\overline{X}-u_{\alpha/2}\cdot\frac{\sigma_0}{\sqrt{n}}<\mu<\overline{X}+u_{\alpha/2}\cdot\frac{\sigma_0}{\sqrt{n}}\right\}=1-\alpha,$$

由此可得参数 μ 的 $1-\alpha-$ 置信区间为

$$\left(\overline{X}-u_{\alpha/2}\cdot\frac{\sigma_0}{\sqrt{n}},\overline{X}+u_{\alpha/2}\cdot\frac{\sigma_0}{\sqrt{n}}\right),$$

从 (7.2.2) 式可知, 其置信区间长度为 $2u_{\alpha/2}\cdot\frac{\sigma_0}{\sqrt{n}}$. 当样本容量 n 固定时, α 减小, 则 $u_{\alpha/2}$ 就增加, 从而区间长度 $2u_{\alpha/2}\cdot\frac{\sigma_0}{\sqrt{n}}$ 也增加.

1. 参数 μ 的区间估计

(2) 设 σ^2 未知. 令

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}},$$

由推论 6.3.1 知 $t \sim t(n-1)$, 对给定的置信水平 $1-\alpha$, 查 $t-$ 分布表可得

$$P\left\{|t| < t_{\alpha/2}(n-1)\right\} = 1 - \alpha,$$

即

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

由此可得参数 μ 的 $1-\alpha-$ 置信区间为

$$\left(\overline{X} - t_{\alpha/2}(n-1) \cdot \frac{S}{\sqrt{n}},\ \overline{X} + t_{\alpha/2}(n-1) \cdot \frac{S}{\sqrt{n}}\right).$$

例 (7.2.1)

设某车间生产一批零件，其直径 $X \sim N(\mu, \sigma^2)$，现从中抽取 9 个，测得数据如下（单位：mm）：

$$19.7, 20.1, 19.9, 19.8, 20.2, 20.1, 20.0, 20.3, 19.0$$

求这批零件直径均值 μ 的置信度为 95% 的置信区间.

解

由于 σ^2 未知, 给定 $\alpha = 0.05$, 根据 (7.2.3) 式可得 μ 的置信度为 95% 的置信区间为

$$\left(\overline{X} - t_{0.025}(n-1)\cdot\frac{S}{\sqrt{n}}, \overline{X} + t_{0.025}(n-1)\cdot\frac{S}{\sqrt{n}}\right),$$

由以上数据可得 $\overline{x} = 19.9, s^2 = 0.15 = (0.387)^2$. 查表 $t_{0.025}(8) = 2.306$. 因此

$$\overline{x} - t_{0.025}(n-1)\cdot\frac{s}{\sqrt{n}} = 19.9 - 2.306 \times \frac{0.387}{3} = 19.603,$$
$$\overline{x} + t_{0.025}(n-1)\cdot\frac{s}{\sqrt{n}} = 19.9 + 2.306 \times \frac{0.387}{3} = 20.197,$$

所以 μ 的一个置信度为 95% 的置信区间为 $(19.603, 20.197)$.

2. 参数 σ^2 的区间估计

设 μ 未知, 根据第一节点估计知, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ 为 σ^2 的无偏估计量. 因此可考虑

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{\sigma^2}$$

由 Fisher 定理 $\chi^2 \sim \chi^2(n-1)$, 给定置信水平 $1-\alpha$, 查 χ^2 分布表可得

$$P\left\{\chi^2_{1-\alpha/2}(n-1) < \chi^2 < \chi^2_{\alpha/2}(n-1)\right\} = 1-\alpha,$$

即

$$P\left\{\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right\} = 1-\alpha.$$

2. 参数 σ^2 的区间估计

因此 σ^2 的置信度为 $1-\alpha-$ 置信区间为

$$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)},\quad \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right).$$

进一步可得总体标准差 σ 的置信度为 $1-\alpha-$ 置信区间为

$$\left(\sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}}, \sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}}\right).$$

例 (7.2.2)

在例 7.2.1 中, 设 μ 未知, 求方差 σ^2 和标准差 σ 的置信度为 0.99 的置信区间.

例 (7.2.2)

在例 7.2.1 中, 设 μ 未知, 求方差 σ^2 和标准差 σ 的置信度为 0.99 的置信区间.

解

由于 μ 未知, 给定 $\alpha = 0.01$, 根据 (7.2.4) 式可得 σ^2 的置信度为 99% 的置信区间为

$$\left(\frac{(n-1)S^2}{\chi^2_{0.005}(n-1)}, \quad \frac{(n-1)S^2}{\chi^2_{0.995}(n-1)}\right)$$

例 7.2.1 中已算得 $s^2 = 0.15$, 因此 $(n-1)S^2 = 8 \times 0.15 = 1.2$, 查 χ^2 分布表得 $\chi^2_{0.005}(8) = 21.955, \chi^2_{0.995}(8) = 1.344$, 由此可得 σ^2 的一个置信区间为

$$\left(\frac{1.2}{21.955}, \frac{1.2}{1.344}\right) = (0.055, 0.893).$$

标准差 σ 的一个置信区间为 $(0.235, 0.945)$.

1 点估计

2 区间估计

- 单个正态总体参数的区间估计
- 两个正态总体参数的区间估计

设总体 $X \sim N\left(\mu_1, \sigma_1^2\right)$, $Y \sim N\left(\mu_2, \sigma_2^2\right)$, 且 X 与 Y 独立, $(X_1, X_2, \cdots, X_n)$, $(Y_1, Y_2, \cdots, Y_m)$ 分别是来自总体 X 与 Y 的样本, 下面讨论参数 $\mu_1 - \mu_2$ 和 σ_1^2/σ_2^2 的区间估计.

1. 均值差 $\mu_1 - \mu_2$ 的区间估计

设方差 σ_1^2 与 σ_2^2 未知但相等，令 $\sigma_1^2 = \sigma_2^2 = \sigma^2$，根据第 6 章知识，可考虑

$$t = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{S_w\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}},$$

由推论 6.3.3 知 $t \sim t(n+m-2)$，对给定的置信水平，查 $t-$ 分布表可得

$$P\left\{|t| < t_{\alpha/2}(n+m-2)\right\} = 1-\alpha,$$

即

$$P\left\{\overline{X} - \overline{Y} - t_{\alpha/2}(n+m-2)\cdot S_w\sqrt{\frac{1}{n}+\frac{1}{m}} < \mu_1 - \mu_2 < \overline{X} - \overline{Y} + t_{\alpha/2}(n+m-2)\cdot S_w\sqrt{\frac{1}{n}+\frac{1}{m}}\right\} = 1-\alpha,$$

1. 均值差 $\mu_1 - \mu_2$ 的区间估计

由此可得参数 $\mu_1 - \mu_2$ 的 $1-\alpha$ − 置信区间为

$$\boxed{\begin{aligned}&\Bigg(\overline{X}-\overline{Y}-t_{\alpha/2}(n+m-2)\cdot S_w\sqrt{\frac{1}{n}+\frac{1}{m}},\\&\overline{X}-\overline{Y}+t_{\alpha/2}(n+m-2)\cdot S_w\sqrt{\frac{1}{n}+\frac{1}{m}}\Bigg).\end{aligned}} \tag{7.2.6}$$

例 (7.2.3)

设两台机床独立地生产同一种钢珠，它们的直径都服从正态分布，且设它们的方差相等. 现从第一台机床生产的钢珠中任意抽取 8 个，从第二台车床生产的钢珠中任意抽取 9 个，测得其直径为 (单位: mm)

$$\begin{aligned}&\text{第一台}: 15.0, 14.8, 15.4, 15.2, 14.9, 15.2, 15.1, 14.8;\\&\text{第二台}: 15.0, 15.2, 15.1, 14.8, 15.0, 14.6, 15.1, 14.5, 14.8,\end{aligned}$$

试求两台车床生产的钢珠直径均值之差的置信度为 95% 的置信区间.

解

X, Y 分别表示第一、第二台车床生产的钢珠直径，于是 $X \sim N\left(\mu_1, \sigma^2\right)$, $Y \sim N\left(\mu_2, \sigma^2\right)$.

解

根据 (7.2.6) 式可得 $\mu_1-\mu_2$ 的置信度为 95% 的置信区间为

$$\left(\overline{X}-\overline{Y}-t_{0.025}(n+m-2)\cdot S_w\sqrt{\frac{1}{n}+\frac{1}{m}},\right.$$

$$\left.\overline{X}-\overline{Y}+t_{0.025}(n+m-2)\cdot S_w\sqrt{\frac{1}{n}+\frac{1}{m}}\right).$$

由以上数据可得 $\overline{x}=15.05, \overline{y}=14.9,$

$$s_w=\sqrt{\left[\sum_{i=1}^{8}(x_i-\overline{x})^2+\sum_{j=1}^{9}(y_j-\overline{y})^2\right]\Big/(n+m-2)}$$

$$=\sqrt{\frac{0.32+0.46}{8+9-2}}=0.228.$$

查 $t-$ 分布表得 $t_{0.025}(15)=2.1315$. 因此 $\mu_1-\mu_2$ 的一个置信区间为 $(-0.862, 0.386)$.

2. 方差比 σ_1^2/σ_2^2 的区间估计

设均值 μ_1, μ_2 未知，要求 σ_1^2/σ_2^2 的置信区间，可考虑

$$F = \frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2},$$

其中 S_1^2, S_2^2 分别为 σ_1^2, σ_2^2 的无偏估计量，根据推论 6.3.3 知 $F \sim F(n-1, m-1)$，给定置信水平 $1-\alpha$，查 $F-$ 分布表得

$$P\left\{F_{1-\alpha/2}(n-1, m-1) < F < F_{\alpha/2}(n-1, m-1)\right\} = 1-\alpha,$$

于是 σ_1^2/σ_2^2 的置信度为 $1-\alpha$ 的置信区间为

$$\left(\frac{S_1^2/S_2^2}{F_{\alpha/2}(n-1, m-1)}, \frac{S_1^2/S_2^2}{F_{1-\alpha/2}(n-1, m-1)}\right).$$