参数估计

李孟棠，朱彬

中山大学智能工程学院

limt29@mail.sysu.edu.cn, zhub26@mail.sysu.edu.cn

2021 年 x 月 x 日

主要内容

本课程只介绍频率学派 (frequentist) 的统计推断方法，主要处理**参数估计**和**假设检验**两类问题。

本章讨论总体（随机变量）分布的参数估计问题，主要分为点估计 (point estimation) 和区间估计 (interval estimation)。

注意频率学派关于未知参数的假设：存在一个确定性的 (deterministic/nonrandom) 真实参数值。

主要内容

1. 点估计
   - 矩估计法
   - 最大似然估计法
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

设总体 X 的分布函数形式已知，但它的一个或多个参数未知。借助总体的样本来估计分布函数中**未知参数值**的问题称为参数的点估计问题。

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 X 服从参数为 $\lambda > 0$ 的 Poisson 分布，λ 未知。现有以下样本值，试估计参数 λ。

着火次数 k	0	1	2	3	4	5	6	≥ 7	
发生 k 次着火的天数 n_k	75	90	54	22	6	2	1	0	$\sum = 250$

解:由于 $X \sim \pi(\lambda)$，故 $\lambda = \mathbb{E}(X)$。很自然地，我们用样本均值来估计总体均值 $\mathbb{E}(X)$。计算得

$$\bar{x} = \frac{\sum_{k=0}^{6} kn_k}{\sum_{k=0}^{6} n_k} = \cdots = 1.22.$$

因此 $\lambda = \mathbb{E}(X)$ 的估计为 1.22。

点估计问题一般表述为：设总体 X 的分布函数 $F(x;\theta)$ 的形式已知，θ 是待估参数。为讨论方便，假设 $\theta \in \mathbb{R}$。$X_1,\ldots,X_n$ 是 X 的一组样本，相应的样本观察值为 $x_1,\ldots,x_n$。构造一个合适的统计量 $\hat{\theta}(X_1,\ldots,X_n)$，用它的观察值 $\hat{\theta}(x_1,\ldots,x_n)$ 作为未知参数 θ 的近似值。

引入名词：

- 称 $\hat{\theta}(X_1,\ldots,X_n)$ 为 θ 的估计器（estimator，书上称“估计量”），是**一个随机变量**，
- 称 $\hat{\theta}(x_1,\ldots,x_n)$ 为 θ 的估计（estimate，书上称“估计值”），是估计器的观察值。

注意：把两组不同的样本 $(x_1,\ldots,x_n)$ 和 $(y_1,\ldots,y_n)$ 放入同一个估计器，一般会得到不同的估计，即 $\hat{\theta}(x_1,\ldots,x_n) \neq \hat{\theta}(y_1,\ldots,y_n)$。例如一般而言，

$$\frac{x_1+\cdots+x_n}{n} \neq \frac{y_1+\cdots+y_n}{n}.$$

在第一个例子中，我们用样本均值来估计总体均值。此时估计器为

$$\hat{\lambda}(X_1, \ldots, X_n) = \widehat{\mathbb{E}(X)} = \frac{1}{n}\sum_{k=1}^{n} X_k, \text{ 其中} n = 250,$$

估计为

$$\hat{\lambda}(x_1, \ldots, x_n) = \frac{1}{n}\sum_{k=1}^{n} x_k = 1.22.$$

主要内容

1. 点估计
   - 矩估计法
   - 最大似然估计法
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

矩估计法

给定总体 X 的样本 $X_1, \ldots, X_n$，基于**样本矩**来构造未知参数的估计。

- 若 X 为连续型随机变量，设其概率密度函数为 $f(x; \theta_1, \ldots, \theta_k)$；
- 若 X 为离散型随机变量，设其概率质量 (probability mass) 函数（即分布律）为 $P(X = x) = p(x; \theta_1, \ldots, \theta_k)$，

其中 $\theta_1, \ldots, \theta_k$ 为待估参数。

假设总体 X 的前 k 阶矩

$$\mu_\ell := \mathbb{E}(X^\ell) = \int_{-\infty}^{\infty} x^\ell f(x; \theta_1, \ldots, \theta_k) \mathrm{d}x \ (X \text{ 为连续型})$$

$$\text{或 } \mu_\ell := \mathbb{E}(X^\ell) = \sum_{x \in R_X} x^\ell p(x; \theta_1, \ldots, \theta_k) \mathrm{d}x \ (X \text{ 为离散型})$$

存在，其中 R_X 表示 X 的值域 (range)，$\ell = 1, 2, \ldots, k$。每个 μ_ℓ 都是参数 $\theta_1, \ldots, \theta_k$ 的函数。

如此得到由 k 个方程组成的矩方程组

$$\begin{cases} \mu_1 & = & g_1(\theta_1, \ldots, \theta_k) \\ \mu_2 & = & g_2(\theta_1, \ldots, \theta_k) \\ & \vdots & \\ \mu_k & = & g_k(\theta_1, \ldots, \theta_k) \end{cases}.$$

一般而言，这是一个包含 k 个未知量的非线性方程组。如果我们**足够幸运**，能够从中解出

$$\begin{cases} \theta_1 & = & h_1(\mu_1, \ldots, \mu_k) \\ \theta_2 & = & h_2(\mu_1, \ldots, \mu_k) \\ & \vdots & \\ \theta_k & = & h_k(\mu_1, \ldots, \mu_k) \end{cases}, \tag{1}$$

那么利用样本矩

$$A_\ell = \frac{1}{n}\sum_{i=1}^{n} X_i^\ell$$

代替(1)中的 μ_ℓ，$\ell = 1, \ldots, k$，就得到了参数 θ_i 的**矩估计器**。

也就是说，对于 $i=1,\ldots,k$，定义估计器

$$\hat{\theta}_i = h_i(A_1,\ldots,A_k),$$

其观察值称为**矩估计**。

采用矩估计器的理由：样本矩依概率收敛到总体矩，即

$$A_\ell \xrightarrow{P} \mu_\ell, \quad \ell = 1,\ldots,k.$$

假设每个 h_i 都是连续函数，那么由**连续映射定理**知

$$h_i(A_1,\ldots,A_k) \xrightarrow{P} h_i(\mu_1,\ldots,\mu_k),$$

即矩估计器 $\hat{\theta}_i \xrightarrow{P} \theta_i$（参数的真实值），$i=1,\ldots,k$。

设总体 X 在 $[a,b]$ 上服从均匀分布，参数 a,b 未知。$X_1,\ldots,X_n$ 是来自 X 的样本，试求 a,b 的矩估计器。

解: 根据均匀分布的定义，稍做计算可得

$$\mu_1=\mathbb{E}(X)=\frac{a+b}{2},$$

$$\mu_2=\mathbb{E}(X^2)=D(X)+[\mathbb{E}(X)]^2=\frac{(b-a)^2}{12}+\frac{(a+b)^2}{4}.$$

整理得

$$\begin{cases} a+b & = & 2\mu_1 \\ b-a & = & \sqrt{12(\mu_2-\mu_1^2)} \end{cases}.$$

解该方程组得到

$$a=\mu_1-\sqrt{3(\mu_2-\mu_1^2)},\quad b=\mu_1+\sqrt{3(\mu_2-\mu_1^2)}.$$

分别用 A_1,A_2 代替 μ_1,μ_2 得到 a 的矩估计器为

$$\hat{a}=A_1-\sqrt{3(A_2-A_1^2)}=\bar{X}-\sqrt{\frac{3}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2},$$

$\hat{b}$ 的表达式只相差一个符号。

例

设总体 X 的均值 μ 及方差 $\sigma^2 > 0$ 都存在但均未知。又设 $X_1, \ldots, X_n$ 是来自 X 的样本。试求 μ, σ^2 的矩估计器。

解:由一、二阶矩和方差的定义知

$$\begin{cases} \mu_1 = \mathbb{E}(X) = \mu \\ \mu_2 = \mathbb{E}(X^2) = D(X) + [\mathbb{E}(X)]^2 = \sigma^2 + \mu^2 \end{cases}.$$

解得

$$\mu = \mu_1, \quad \sigma^2 = \mu_2 - \mu_1^2.$$

分别以 A_1, A_2 代替 μ_1, μ_2，得到 μ 和 σ^2 的矩估计器分别为

$$\begin{aligned} \hat{\mu} &= A_1 = \bar{X}, \\ \widehat{\sigma^2} &= A_2 - A_1^2 = \frac{1}{n} X_i^2 - \bar{X}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2. \end{aligned} \tag{2}$$

上述结果表明，总体均值和方差的矩估计器和总体分布的具体形式无关。例如，若 $X \sim N(\mu, \sigma^2)$，参数 μ, σ^2 未知，那么它们的矩估计器由(2)给出。

主要内容

最大似然 (Maximum Likelihood) 估计法

如果总体 X 属于离散型，假设其概率质量函数

$$P\{X = x\} = p(x;\theta), \quad \theta \in \Theta \subset \mathbb{R}$$

的形式已知，其中 θ 为待估参数，Θ 为 θ 可能取值的范围。

设 $X_1, \ldots, X_n$ 是来自 X 的样本，则它们的联合分布律为

$$P\{X_1 = x_1, \ldots, X_n = x_n\} = \prod_{i=1}^{n} p(x_i;\theta).$$

给定样本的观察值 $x_1, \ldots, x_n$，定义参数 θ 的**似然函数** (likelihood function)

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

其中 $x_1, \ldots, x_n$ 是给定的样本值。

最大似然估计

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \ L(\theta; x_1, \ldots, x_n) \tag{3}$$

也写作 $\hat{\theta} = \hat{\theta}(x_1, \ldots, x_n)$。相应的统计量 $\hat{\theta}(X_1, \ldots, X_n)$ 称为参数 θ 的最大似然估计器。

原理：直观上而言，最大似然法选择最优参数 $\hat{\theta}$ 使得**给定的样本观察值最有可能被取到**，也就是最大化参数对数据的“解释力度”或者数据对参数的“支持程度”。

类似地，如果总体 X 属于连续型，设其概率密度 $f(x; \theta)$, $\theta \in \Theta$ 的形式已知。给定样本 $X_1, \ldots, X_n$，它们的联合概率密度为 $\prod_{i=1}^{n} f(x_i; \theta)$。设样本观察值为 $x_1, \ldots, x_n$，定义似然函数

$$L(\theta) = L(\theta; x_1, \ldots, x_n) := \prod_{i=1}^{n} f(x_i; \theta), \quad \theta \in \Theta.$$

相应的最大似然估计仍然由(3)给出。

所以，求解最大似然估计问题就转化为微积分中的函数极值问题。

在很多情形下，概率质量函数 $p(x;\theta)$ 或密度函数 $f(x;\theta)$ 关于参数 θ 可微，此时 $\hat{\theta}$ 必满足驻点方程（一阶必要条件，但**不充分**）

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

由于 $\ln$ 函数的单调性，$L(\theta)$ 和 $\ln L(\theta)$ 的极值点相同，所以最大似然估计 $\hat{\theta}$ 也可以从**对数似然方程**

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

解得。

通常，我们处理对数似然 (log-likelihood) 函数更为方便。

例

设 $X \sim b(1,p)$，即参数为 p 的 $(0-1)$ 分布，$X_1,\ldots,X_n$ 是来自 X 的样本。试求参数 p 的最大似然估计器。

解:X 的分布律为

X	0	1
$P\{X=x\}$	$1-p$	p

设 $x_1,\ldots,x_n$ 是样本观察值，则似然函数为

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

而对数似然函数为

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

令

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

解得 p 的最大似然估计 $\hat{p}=\frac{1}{n}\sum_{i=1}^n x_i=\bar{x}$，相应的估计器为 $\hat{p}=\bar{X}$，与矩估计器相同。

最大似然法也适用于总体分布中含有多个未知参数 $\theta_1, \ldots, \theta_k$ 的情形，此时分别令

$$\frac{\partial}{\partial \theta_i} L(\theta_1, \ldots, \theta_k), \quad i = 1, \ldots, k,$$

或

$$\frac{\partial}{\partial \theta_i} \ln L(\theta_1, \ldots, \theta_k), \quad i = 1, \ldots, k. \tag{4}$$

求解上述包含 k 个未知量的 k 个方程，即可得到未知参数的最大似然估计。式(4)称为**对数似然方程组**。

一般而言，最大似然问题是一个关于多个参数 $\theta_1, \ldots, \theta_k$ 的**非线性**、**非凸**优化问题，往往没有解析解，而需要采用数值方法求解（超出本课程范围）。在许多实际问题情境，最大似然估计是一个富有挑战性的问题。

例

设总体 $X \sim N(\mu, \sigma^2)$，参数 μ, σ^2 未知。$x_1, \ldots, x_n$ 是来自 X 的一组样本值。求 μ, σ^2 的最大似然估计。

解:X 的概率密度为

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

似然函数为

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

对应的对数似然函数为

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

例 (续)

令

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

解得 $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$，$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ 因此 μ 和 σ^2 的最大似然估计器分别为

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

再一次与矩估计器相同。

例

设总体 X 在 $[a,b]$ 上服从均匀分布，其中参数 a,b 未知。$x_1,\ldots,x_n$ 为一组样本值。试求 a,b 的最大似然估计。

解：记 $x_{\min} := \min\{x_1,\ldots,x_n\}$，$x_{\max} := \max\{x_1,\ldots,x_n\}$。$X$ 的概率密度为

$$f(x;a,b)=\begin{cases}\dfrac{1}{b-a}, & a\le x\le b,\\[2ex] 0, & \text{其他}.\end{cases}$$

似然函数为

$$L(a,b)=\begin{cases}\dfrac{1}{(b-a)^n}, & a\le x_1,\ldots,x_n\le b,\\[2ex] 0, & \text{其他},\end{cases}$$

其中第一行条件等价于 $a\le x_{\min}$ 和 $x_{\max}\le b$，在此条件下，显然有

$$L(a,b)=\frac{1}{(b-a)^n}\le\frac{1}{(x_{\max}-x_{\min})^n}.$$

等号在 $a=x_{\min}$，$b=x_{\max}$ 时成立，即得到似然函数的最大值。

因此参数 a, b 的最大似然估计器为

$$\hat{a} = \min_{1 \le i \le n} X_i, \quad \hat{b} = \max_{1 \le i \le n} X_i.$$

注意，此处的解与矩估计器**不同**。

Proposition (最大似然估计的函数不变性 (functional invariance))

如果 $\hat{\theta}$ 是参数 θ 的最大似然估计，$g(\theta)$ 是 θ 的任意函数（不需要可逆），那么参数 $u = g(\theta)$ 的最大似然估计为 $\hat{u} = g(\hat{\theta})$。

证明.

这里仅在函数 $g : \Theta \to \mathscr{U}$ 可逆的额外假设下进行证明。令 $h = g^{-1} : \mathscr{U} \to \Theta$。已知 $\hat{\theta} = h(\hat{u})$，且

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \; L(\theta; x_1, \ldots, x_n).$$

因此，根据定义有

$$\hat{u} = \underset{u \in \mathscr{U}}{\operatorname{argmax}} \; L(h(u); x_1, \ldots, x_n).$$

□

函数不变性的应用:

在前面的一个例题中，我们知道正态总体的方差 σ^2 的最大似然估计为

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

取 $g(\theta) = \sqrt{\theta}$。由此得到标准差 $\sigma = g(\sigma^2)$ 的最大似然估计为

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

主要内容

研究某个产品的可靠性时，需要考察其寿命 T。这是个随机变量，为了对其分布进行统计推断，需要通过实验获取产品的寿命数据。

一个典型的寿命实验：将随机抽取的 n 个产品在时刻 $t=0$ 时同时投入实验，直到每个产品都失效，记录每个产品的失效时间。这样得到的样本（即所有产品的失效时间 $0 \le t_1 \le t_2 \le \cdots \le t_n$）称为**完全样本**。

然而由于时间和财力的限制，我们很难得到完全样本。于是转而考虑**截尾寿命实验**，主要分为以下两种。

- **定时截尾样本**：随机抽取 n 个产品在 $t=0$ 时同时投入实验，进行到事先规定的截尾时间 t_0 停止。如果停止时共有 m 个产品失效，失效时间（样本）分别为

$$0 \le t_1 \le t_2 \le \cdots \le t_m \le t_0.$$

  此时 m 是一个随机变量。

- **定数截尾样本**: 随机抽取 n 个产品在 $t=0$ 时同时投入实验，进行到有 m 个产品失效时停止，其中 $m<n$ 为事先规定的数字。m 个失效产品的失效时间（样本）分别为

$$0 \le t_1 \le t_2 \le \cdots \le t_m.$$

此时 t_m（实际上，所有的 t_i）是随机变量。

下面考虑用截尾样本进行统计推断。设产品的寿命 T 服从指数分布，其概率密度为

$$f(t)=\begin{cases} \frac{1}{\theta}e^{-t/\theta}, & t>0, \\ 0, & t\le 0, \end{cases}$$

其中 $\theta>0$ 为待估参数（注意 $\mathbb{E}(T)=\theta$）。

我们用最大似然法估计 θ。首先考虑上面的定数截尾样本。

在时间区间 $[0, t_m]$ 中有 m 个产品失效，其余 $n-m$ 个产品在 t_m 时刻尚未失效（它们的寿命超过 t_m）。考虑这些事件的概率以确定似然函数。

- 一个产品在 $(t_i, t_i + dt_i]$ 失效的概率近似为 $f(t_i)\mathrm{d}t_i = \frac{1}{\theta} e^{-t_i/\theta}\mathrm{d}t_i$，$i = 1, \ldots, m$。
- 一个产品寿命超过 t_m 的概率为 $\int_{t_m}^{\infty} \frac{1}{\theta} e^{-t/\theta}\mathrm{d}t = e^{-t_m/\theta}$。

所以上述观察结果出现的概率近似为

$$\binom{n}{m}\left(\frac{1}{\theta}e^{-t_1/\theta}\mathrm{d}t_1\right)\left(\frac{1}{\theta}e^{-t_2/\theta}\mathrm{d}t_2\right)\cdots\left(\frac{1}{\theta}e^{-t_m/\theta}\mathrm{d}t_m\right)\left(e^{-t_m/\theta}\right)^{n-m}$$

$$=\binom{n}{m}\underbrace{\frac{1}{\theta^m}e^{-\frac{1}{\theta}[t_1+t_2+\cdots+t_m+(n-m)t_m]}}_{\text{取为似然函数}L(\theta)}\mathrm{d}t_1\mathrm{d}t_2\cdots\mathrm{d}t_m,$$

其中 $\mathrm{d}t_1, \ldots, \mathrm{d}t_m$ 为无穷小“常量”。

得到对数似然函数为

$$\ln L(\theta) = -m \ln \theta - \frac{1}{\theta} \underbrace{\left[t_1 + t_2 + \cdots + t_m + (n-m) t_m\right]}_{:= s(t_1,\ldots,t_m)}.$$

令

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \ln L(\theta) = -\frac{m}{\theta} + \frac{1}{\theta^2} s(t_1, \ldots, t_m) = 0,$$

解得 θ 的最大似然估计为

$$\hat{\theta} = \frac{s(t_1, \ldots, t_m)}{m},$$

其中 $s(t_1, \ldots, t_m)$ 称为**总实验时间**，即直到时刻 t_m 为止，n 个产品试验时间的总和。

对于以 t_0 为截止时间的定时截尾样本

$$0 \le t_1 \le t_2 \le \cdots \le t_m \le t_0,$$

我们可做类似分析，得到似然函数为

$$L(\theta) = \frac{1}{\theta^m} e^{-\frac{1}{\theta}[t_1 + t_2 + \cdots + t_m + (n-m)t_0]}.$$

此时参数 θ 的最大似然估计为

$$\hat{\theta} = \frac{s(t_1, \ldots, t_m, t_0)}{m},$$

其中 $s(t_1, \ldots, t_m, t_0) := t_1 + t_2 + \cdots + t_m + (n-m)t_0$ 也称为总实验时间，即直到时刻 t_0 为止，n 个产品试验时间的总和。

例

设一批电池的寿命服从指数分布，其概率密度为

$$f(t) = \begin{cases} \frac{1}{\theta} e^{-t/\theta}, & t > 0, \\ 0, & t \le 0, \end{cases}$$

其中参数 $\theta > 0$ 未知。随机取 50 只电池投入寿命实验，规定实验进行到有 15 只电池失效时结束。测得失效时间（以 h 计）为

115	119	131	138	142	147	148	155
158	159	163	166	167	170	172	

试求电池的平均寿命 θ 的最大似然估计。

解: 这是一组定数截尾样本。对应前面的分析，我们有 $n = 50$，$m = 15$，

$$s(t_1, \ldots, t_m) = \cdots = 8270.$$

所以 θ 的最大似然估计为

$$\hat{\theta} = \frac{s(t_1, \ldots, t_m)}{m} = \frac{8270}{15} \approx 551.33 \text{ (h)}.$$

主要内容

对于同一参数，不同估计方法求出的估计器可能不同。例如在均匀总体的参数估计问题中，矩估计与最大似然估计的结果不同。

原则上，任何统计量都可以作为未知参数的估计。如何评价一个估计器的优劣，标准有哪些？

常用的评价标准包括：无偏性 (unbiasedness)、有效性 (efficiency)、一致性（consistency，书上称相合性）。

无偏性

设 $X_1, \ldots, X_n$ 是总体 X 的一组样本，$\theta \in \Theta$ 是总体分布中的未知参数，其中 $\Theta \subset \mathbb{R}$ 是参数的取值范围。

如果估计器 $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ 的数学期望 $\mathbb{E}(\hat{\theta})$ 存在，且对于任意 $\theta \in \Theta$ 有

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

则称 $\hat{\theta}$ 是 θ 的**无偏估计器**。

含义：对于某些样本观察值 $(x_1, \ldots, x_n)$，$\hat{\theta}$ 的观察值相对于真值来说偏大，有些则偏小。在不同的样本观察值上反复使用估计器 θ，把得到的结果“平均”之后偏差（近似）为零。

在科学技术中 $\mathbb{E}(\hat{\theta}) - \theta$ 常被称为估计器 $\hat{\theta}$ 的**系统误差**，所以无偏估计器满足系统误差为零。

例如，设总体 X 的均值 μ，方差 $\sigma^2>0$ 均未知，由书本第六章 pp. 145–146 知

$$\mathbb{E}(\bar{X})=\mu,\quad \mathbb{E}(S^2)=\sigma^2.$$

也就是说，无论总体服从什么分布，样本均值 $\bar{X}$ 是总体均值 μ 的无偏估计；样本方差 $S^2=\frac{1}{n-1}\sum_{i=1}^n(X_i-\bar{X})^2$ 是总体方差的无偏估计。

因此，估计器 $\frac{1}{n}\sum_{i=1}^n(X_i-\bar{X})^2$ **不是** σ^2 的无偏估计。一般我们更倾向于取 S^2 作为 σ^2 的估计器。

例

设总体 X 的 k 阶矩 $\mu_k=\mathbb{E}(X^k)$ 存在（$k\geq 1$）。又设 $X_1,\ldots,X_n$ 是 X 的一组样本。试证明：无论总体服从什么分布，样本 k 阶矩 $A_k=\frac{1}{n}\sum_{i=1}^n X_i^k$ 是总体 k 阶矩 μ_k 的无偏估计器。

解：因为 $X_1,\ldots,X_n$ 和 X 同分布，所以

$$\mathbb{E}(X_i^k)=\mathbb{E}(X^k)=\mu_k,\quad i=1,\ldots,n.$$

因此

$$\mathbb{E}(A_k)=\frac{1}{n}\sum_{i=1}^n\mathbb{E}(X_i^k)=\mu_k.$$

例 (2)

设总体 X 服从指数分布 $\mathrm{Exp}(\theta)$，其概率密度为

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

其中参数 $\theta>0$ 未知。又设 $X_1,\ldots,X_n$ 是 X 的一组样本。试证:$\bar{X}$ 和 $nZ=n(\min\{X_1,\ldots,X_n\})$ 都是 θ 的无偏估计器。

解: 因为 $\mathbb{E}(\bar{X})=\mathbb{E}(X)=\theta$，所以 $\bar{X}$ 是 θ 的无偏估计器。
而随机变量 $Z=\min\{X_1,\ldots,X_n\}$ 的分布函数为 (书本 p. 83)

$$F_{\min}(z)=1-[1-F(z)]^n,\quad \text{其中}\quad F(x)=\begin{cases}1-e^{-x/\theta}, & x>0,\\ 0, & x\le 0,\end{cases}$$

为 X 的分布函数。等式两边求导得到 Z 的概率密度为

$$f_{\min}(z)=F'_{\min}(z)=nf(z)[1-F(z)]^{n-1}=\begin{cases}\frac{n}{\theta}e^{-nz/\theta}, & z>0,\\ 0, & z\le 0,\end{cases}$$

也就是 $Z\sim\mathrm{Exp}(\theta/n)$。因此，$\mathbb{E}(Z)=\theta/n$，$\mathbb{E}(nZ)=\theta$，无偏性得证。

有效性

对于固定的样本容量 n，比较参数 θ 的两个无偏估计器 $\hat{\theta}_1$ 和 $\hat{\theta}_2$。如果 $\hat{\theta}_1$ 的观察值比 $\hat{\theta}_2$ 在真值 θ 附近更为“密集”，我们就认为 $\hat{\theta}_1$ 更好。

对于无偏估计器 $\hat{\theta}$（一个随机变量），方差 $D(\hat{\theta})$ 度量了它的取值和它数学期望 $\mathbb{E}(X) = \theta$ 之间的偏离程度，因此自然 $D(\hat{\theta})$ 越小越好。

定义 (有效性)

设 $\hat{\theta}_j = \hat{\theta}_j(X_1, \ldots, X_n)$ 是 θ 的无偏估计器，$j = 1, 2$。如果对于任意 $\theta \in \Theta$，有

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

且至少对于某一个 $\theta \in \Theta$ 上式中的不等号严格成立，则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 有效。

例 (2 续)

设总体 X 服从指数分布 $\mathrm{Exp}(\theta)$，其概率密度为

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

其中参数 $\theta>0$ 未知。又设 $X_1,\ldots,X_n$ 是 X 的一组样本，由前面的计算得知 $\bar{X}$ 和 $nZ=n(\min\{X_1,\ldots,X_n\})$ 都是 θ 的无偏估计器。证明：当 $n>1$ 时，$\bar{X}$ 比 nZ 有效。

解：由于 $D(X)=\theta^2$，故 $D(\bar{X})=\theta^2/n$。类似地，由于 $D(Z)=\theta^2/n^2$，故 $D(nZ)=\theta^2$。易知当 $n>1$ 时，$D(\bar{X})<D(nZ)$。因此 $\bar{X}$ 比 nZ 有效。

一致性

讨论估计器的无偏性和有效性时，我们假设样本容量 n 固定。

随着样本容量的增大，我们希望估计器的（观察）值稳定于待估参数的真值。

定义 (一致性)

设 $\hat{\theta}(X_1, \ldots, X_n)$ 为参数 θ 的一个估计器，若对于任意 $\theta \in \Theta$，当 $n \to \infty$ 时 $\hat{\theta}(X_1, \ldots, X_n)$ 依概率（或几乎肯定）收敛于 θ，也就是说，对于任意 $\theta \in \Theta$，都有

$$\forall \varepsilon > 0: \lim_{n\to\infty} P\left\{|\hat{\theta} - \theta| < \varepsilon\right\} = 1,$$

或

$$P\left\{\lim_{n\to\infty} \hat{\theta} = \theta\right\} = 1,$$

则称 $\hat{\theta}$ 为 θ 的一致估计器。

一致性是对估计器的基本要求。不具有一致性的估计器通常不可取，因为不论样本容量 n 取多大，都不能将参数 θ 估计得足够准确。

例如在 7.1 节讲述的**矩估计法**中，假设待估参数可以表示为

$$\theta = g(\mu_1, \mu_2, \ldots, \mu_k),$$

其中，整数 $k \geq 1$，$\mu_k = \mathbb{E}(X^k)$ 为总体 X 的 k 阶矩，g 为连续函数。

构造矩估计器

$$\hat{\theta} = g(A_1, A_2, \ldots, A_k),$$

其中 A_k 为样本 k 阶矩。

由大数定律知 $A_k \xrightarrow{P} \mu_k$，再由连续映射定理知 $\hat{\theta} \xrightarrow{P} \theta$，故矩估计器具有一致性。

再看**最大似然估计**，为什么采用这种方法？

- 其历史可以追溯到 Gauss、Laplace 等人；
- 在 1912 到 1922 年期间由 Fisher 在学界普及；
- 性质的严格证明则由美国统计学家 Samuel Wilks 在 1938 年完成。

在 i.i.d. 样本假设下，最大似然估计器具有如下性质：

1. 渐近无偏
$$\lim_{N\to\infty} \mathbb{E}[\hat{\boldsymbol{\theta}}_{\mathsf{ML}}] = \boldsymbol{\theta},$$
其中 $\boldsymbol{\theta}$ 为参数真值。
2. 渐近一致，给定任意 $\varepsilon > 0$，有
$$\lim_{N\to\infty} P\left\{|\hat{\boldsymbol{\theta}}_{\mathsf{ML}} - \boldsymbol{\theta}| > \varepsilon\right\} = 0,$$
也就是依概率收敛到参数真值。
3. 渐近有效，也就是其方差达到 Cramér-Rao 下限。

例 (样本均值 $\bar{X}$ 和样本方差 S^2 的一致性)

样本均值 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 的一致性由大数定律导出。

再考察高斯总体 $X \sim N(\mu, \sigma^2)$ 方差 σ^2 的最大似然估计:

$$\begin{aligned}\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 &= \frac{1}{n}\left(\sum_i X_i^2 - n\bar{X}^2\right) \\ &= \frac{1}{n}\sum_i X_i^2 - \bar{X}^2 \\ &\to \mathbb{E}(X^2) - \mu^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2.\end{aligned}$$

显然

$$S^2 = \frac{n}{n-1}\widehat{\sigma^2} \to 1 \times \sigma^2 = \sigma^2.$$

所以样本方差 S^2 和有偏估计 $\widehat{\sigma^2}$ 都具有一致性。

主要内容

1. 点估计
   - 矩估计法
   - 最大似然估计法
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

置信区间

点估计给出未知参数的一个近似值，**区间估计**给出未知参数的取值范围（通常为一个区间，称为**置信区间**，confidence interval），以及这个范围包含真实参数的可信程度。两类估计的取舍取决于实际应用场景。

Definition (置信区间)

设总体 X 的分布函数 $F(x;\theta)$ 含有一个未知参数 $\theta \in \Theta$，$X_1, X_2, \ldots, X_n$ 是来自 X 的样本。对于给定值 $\alpha \in (0,1)$，如果两个统计量 $\underline{\theta}(X_1, \ldots, X_n)$ 和 $\overline{\theta}(X_1, \ldots, X_n)$ $(\underline{\theta} < \overline{\theta})$ 对于任意 $\theta \in \Theta$ 满足

$$P\left\{\underline{\theta}(X_1, \ldots, X_n) < \theta < \overline{\theta}(X_1, \ldots, X_n)\right\} \geq 1 - \alpha,$$

则称随机区间 $(\underline{\theta}, \overline{\theta})$ 是参数 θ 以 $1-\alpha$ 为**置信水平**的置信区间，$\underline{\theta}$ 和 $\overline{\theta}$ 分别称为置信下限和置信上限。

给定 α 时置信区间的计算：

- 对于连续型随机变量 X，我们总是按等式

$$P\{\underline{\theta} < \theta < \overline{\theta}\} = 1 - \alpha \tag{5}$$

  求置信区间。

- 对于离散型随机变量 X，常常无法找到区间 $(\underline{\theta}, \overline{\theta})$ 使得上述等式严格成立。此时我们寻找 $(\underline{\theta}, \overline{\theta})$ 使得概率 $P\{\underline{\theta} < \theta < \overline{\theta}\}$ 尽可能接近 $1-\alpha$。

原因：我们希望置信区间尽可能窄，使得估计误差得到控制。

等式(5)的相对频率解读为：若固定样本容量 n 做反复抽样，每个样本观察值确定一个区间 $(\underline{\theta}, \overline{\theta})$。那么每个这样的区间包含真实参数 θ 这一事件可能发生或不发生。根据伯努利大数定律，重复抽样构造的多个区间中，包含 θ 真值的约占 $100(1-\alpha)\%$。

例如取 $\alpha = 0.01$，反复抽样 1000 次，那么构造的 1000 的区间中不包含 θ 真值的约有 10 个。

例

设总体 $X \sim N(\mu, \sigma^2)$，其中参数 σ^2 已知，μ 未知。又设 $X_1, \ldots, X_n$ 是来自 X 的样本。求 μ 的置信水平为 $1-\alpha$ 的置信区间。

解：我们知道样本均值 $\bar{X}$ 是 μ 的无偏估计，且有

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

显然上面的分布不依赖于未知参数。按照标准正态分布上 α 分位数的定义，有

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

即

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

这样我们就得到了 μ 的一个置信水平为 $1-\alpha$ 的置信区间

$$\left(\bar{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\,,\ \bar{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right)，\text{常记作 } \left(\bar{X}\pm\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right).$$

如果取 $1-\alpha=0.95$ 即 $\alpha=0.05$，同时给定 $\sigma=1$，$n=16$，查表得 $z_{\alpha/2}=z_{0.025}=1.96$，那么我们得到一个置信水平为 0.95 的置信区间

$$\left(\bar{X}\pm\frac{1}{\sqrt{16}}\times 1.96\right) \text{ 即 } (\bar{X}\pm 0.49)\,. \tag{6}$$

然后，如果由一组样本值计算得到 $\bar{x}=5.2$，则得到一个（非随机）区间

$$(5.2\pm 0.49) \text{ 即 } (4.71, 5.69).$$

我们仍称它为 μ 的置信水平为 0.95 的置信区间。

原因：若固定样本容量 $n=16$ 反复抽样多次，每组样本观察值按(6)确定一个区间。按照概率的相对频率解读，这些区间中包含 μ 真值的约占 95%。因此上面的区间 $(4.71, 5.69)$ 包含 μ 的“可信程度”为 95%。

注意：给定 $\alpha \in (0,1)$，置信水平为 $1-\alpha$ 的置信区间并不唯一。

在上面的例子中，给定 $\alpha = 0.05$，我们可以选取标准正态分布的分位数使得

$$P\left\{-z_{0.04} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{0.01}\right\} = 1-\alpha,$$

即

$$P\left\{\bar{X} - \frac{\sigma}{\sqrt{n}} z_{0.01} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} z_{0.04}\right\} = 1-\alpha.$$

所以

$$\left(\bar{X} - \frac{\sigma}{\sqrt{n}} z_{0.01}\,,\ \bar{X} + \frac{\sigma}{\sqrt{n}} z_{0.04}\right) \tag{7}$$

也是 μ 的置信水平为 0.95 的置信区间。

式(7)中的区间长度为 $\frac{\sigma}{\sqrt{n}}(z_{0.04}+z_{0.01}) \approx 4.08\frac{\sigma}{\sqrt{n}}$，长于前面以“对称”方式构造的置信区间长度 $\frac{\sigma}{\sqrt{n}} \times 2z_{0.25} \approx 3.92\frac{\sigma}{\sqrt{n}}$。因此后者更优。

易知，如果概率密度类似 $N(0,1)$ 分布那样满足单峰且对称（例如 t 分布），那么以“对称”方式构造的置信区间长度最短。

总结构造未知参数 θ 的置信区间的具体做法：

1. 寻求一个样本 $X_1,\ldots,X_n$ 和参数 θ 的函数 $W(X_1,\ldots,X_n;\theta)$，使得它的分布不依赖于 θ 和其他未知参数。具有这种性质的函数 W 称为枢轴量（pivotal quantity，也可译为“关键量”）。

2. 对于给定的置信水平 $1-\alpha$，定出两个常数 a,b 使得

$$P\{a < W(X_1,\ldots,X_n;\theta) < b\} = 1-\alpha.$$

   如果能从不等式 $a < W(X_1,\ldots,X_n;\theta) < b$ 解出未知参数 θ 满足的不等式 $\underline{\theta} < \theta < \overline{\theta}$，其中 $\underline{\theta}(X_1,\ldots,X_n)$ 和 $\overline{\theta}(X_1,\ldots,X_n)$ 都是统计量，那么区间 $(\underline{\theta},\overline{\theta})$ 就是 θ 以 $1-\alpha$ 为置信水平的置信区间。

枢轴量 $W(X_1,\ldots,X_n;\theta)$ 的构造通常从参数 θ 的点估计着手考虑。

主要内容

单个正态总体 $N(\mu,\sigma^2)$ 的情况

给定置信水平 $1-\alpha$。设 $X_1,\ldots,X_n$ 为总体 $N(\mu,\sigma^2)$ 的样本，$\bar{X}$ 和 S^2 分别为样本均值和样本方差。

1. 均值 μ 的置信区间
(1) σ^2 已知。由 7.4 节例子，采用枢轴量 $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$，得到 μ 的置信区间为

$$\left(\bar{X}\pm\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right).$$

(2) σ^2 未知。此时上式不适用，因为它包含了未知参数 σ。考虑到 S^2 是 σ^2 的无偏估计，将上面枢轴量中的 σ 替换为 $S=\sqrt{S^2}$。由 6.3 节定理 4 知，

$$\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1),$$

且右边的分布不依赖于任何未知参数。

因此可使用 $\frac{\bar{X}-\mu}{S/\sqrt{n}}$ 作为枢轴量，得到

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

即

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

于是得到 μ 的置信区间

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

在实际问题中，总体方差 σ^2 未知的情况居多，故上式更具实用价值。

例

有一大批糖果。现从中随机抽取 16 袋，称得质量（以 g 计）如下：

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

设袋装糖果的质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 $1-\alpha=0.95$，$\alpha/2=0.025$，$n-1=15$。查表得 $t_{0.025}(15)=2.1315$。由给出的数据算得 $\bar{x}=503.75$，$s=6.2022$。根据前面的分析，我们计算得到 μ 的一个置信水平为 0.95 的置信区间为

$$\left(503.75 \pm \frac{6.2022}{\sqrt{16}} \times 2.1315\right) \approx (500.4, 507.1).$$

也就是说，这批袋装糖果质量的总体均值在 500.4 g 和 507.1 g 之间的可信程度为 95%。如果以样本均值 $\bar{x}$（即置信区间中点）作为 μ 的近似值，其误差不超过 $\frac{6.2022}{\sqrt{16}} \times 2.1315 \approx 3.305$ g（即半区间长）。这个误差估计的可信程度为 95%。

2. 方差 σ^2 的置信区间。此处根据实际问题的需要，只介绍 μ 未知的情况。

由 6.3 节定理 3 知

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1),$$

且上式右端的分布不依赖于任何未知参数。故取 $\frac{(n-1)S^2}{\sigma^2}$ 作为枢轴量，得到

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

即

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

如此得到 σ^2 的一个置信水平为 $1-\alpha$ 的置信区间

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

顺便得到标准差 σ 的置信区间为

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

注意，当概率密度函数不对称时，如 χ^2 分布和 F 分布，习惯上仍然取对称的分位数对确定置信区间。原则上，取不对称的分位数可以使置信区间更短。

例

求前面例题中总体标准差 σ 的置信水平为 0.95 的置信区间。

解: 现有 $\alpha/2 = 0.025$，$1-\alpha/2 = 0.975$，$n-1 = 15$。查表得 $\chi^2_{0.025}(15) = 27.488$，$\chi^2_{0.975}(15) = 6.262$。前面算得 $s = 6.2022$。代入上面公式，得到标准差 σ 的一个置信水平为 0.95 的置信区间为

$$(4.58, 9.60).$$

两个总体 $N(\mu_1, \sigma_1^2)$，$N(\mu_2, \sigma_2^2)$ 的情况

在实际中我们常遇到以下问题：已知产品的某一质量指标服从正态分布，但由于原料、设备条件、操作人员不同或工艺过程改变等因素，引起总体均值和方差有所改变。

我们需要知道这些变化有多大，此时要考虑两个正态总体均值差或方差比的估计问题。

设给定置信水平 $1-\alpha$。并设 $X_1, \ldots, X_{n_1}$ 是来自第一个总体的样本，$Y_1, \ldots, Y_{n_2}$ 是来自第二个总体的样本，且两组样本相互独立。

设 $\bar{X}, \bar{Y}$ 分别是第一、第二个总体的样本均值，S_1^2, S_2^2 分别是第一、第二个总体的样本方差。

1. 两个总体均值差 $\mu_1 - \mu_2$ 的置信区间

(1) σ_1^2, σ_2^2 均为已知。

- $\bar{X}, \bar{Y}$ 分别是 μ_1, μ_2 的无偏估计 $\Longrightarrow$ $\bar{X} - \bar{Y}$ 是 $\mu_1 - \mu_2$ 的无偏估计。
- 由 $\bar{X} \sim N(\mu_1, \sigma_1^2/n_1)$，$\bar{Y} \sim N(\mu_2, \sigma_2^2/n_2)$ 以及 $\bar{X}, \bar{Y}$ 的独立性知

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right),$$

或

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1).$$

取上式左边为枢轴量，即得到 $\mu_1 - \mu_2$ 的置信区间为

$$\left(\bar{X} - \bar{Y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right).$$

(2) $\sigma_1^2=\sigma_2^2=\sigma^2$，但 σ^2 未知。此时由 6.3 节定理 5 知

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_W\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2),$$

取上式左边为枢轴量，可得 $\mu_1-\mu_2$ 的置信区间为

$$\left(\bar{X}-\bar{Y}\pm t_{\alpha/2}(n_1+n_2-2)S_W\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$$

其中

$$S_W^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2},\quad S_W=\sqrt{S_W^2}.$$

例

为比较 I、II 两种型号步枪子弹的枪口速度，随机取 I 型子弹 10 发，得到枪口速度的平均值为 $\bar{x}_1 = 500$ m/s，标准差 $s_1 = 1.10$ m/s；随机取 II 型子弹 20 发，得到枪口速度的平均值为 $\bar{x}_2 = 496$ m/s，标准差 $s_2 = 1.20$ m/s。假设两总体都近似服从正态分布，且由生产过程可认为方差相等。求两总体均值差 $\mu_1 - \mu_2$ 的一个置信水平为 0.95 的置信区间。

解：按实际情况，可假设来自两个总体的样本相互独立。由于两总体的方差相等且未知，故可采用第二种方法求均值差的置信区间。这里 $1-\alpha = 0.95$，$\alpha/2 = 0.025$，$n_1 = 10$，$n_2 = 20$，$n_1 + n_2 - 2 = 28$，查表得 $t_{0.025}(28) = 2.0484$，$s_w^2 = (9 \times 1.1^2 + 19 \times 1.2^2)/28$，$s_w = \sqrt{s_w^2} = 1.1688$。因此所求置信区间为

$$\left(\bar{x}_1 - \bar{x}_2 \pm t_{0.025}(28) \times s_w \sqrt{\frac{1}{10} + \frac{1}{20}}\right) = (4 \pm 0.93),$$

即 $(3.07, 4.93)$。该区间的下限大于零，在实际中我们就认为 $\mu_1 > \mu_2$。

例

为提高某一化学生产过程的得率，工厂试图采用一种新的催化剂。为慎重起见，首先对此进行实验。设采用原来的催化剂进行了 $n_1 = 8$ 次试验，得到得率的平均值 $\bar{x}_1 = 91.73$，样本方差 $s_1^2 = 3.89$；又采用新的催化剂进行了 $n_2 = 8$ 次试验，得率的平均值为 $\bar{x}_2 = 93.75$，样本方差 $s_2^2 = 4.02$。假设两总体都服从正态分布且方差相等，两组样本独立。试求两总体均值差 $\mu_1 - \mu_2$ 的置信水平为 0.95 的置信区间。

解: 采用和上一题一样的做法，最终所求置信区间为 $(-4.15, 0.11)$。由于该区间包含零，在实际中我们就认为采用这两种催化剂得率的均值没有显著差别（即差值“近似”为零）。

2. 两个总体方差比 σ_1^2/σ_2^2 的置信区间
仅讨论总体均值 μ_1, μ_2 均为未知的情况。由 6.3 节定理 5 知

$$\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1-1, n_2-1),$$

其中右端分布不依赖任何未知参数。因此取 $\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}$ 为枢轴量得

$$P\left\{F_{1-\alpha/2}(n_1-1, n_2-1) < \frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} < F_{\alpha/2}(n_1-1, n_2-1)\right\} = 1-\alpha,$$

即

$$P\left\{\frac{S_1^2}{S_2^2}\frac{1}{F_{\alpha/2}(n_1-1, n_2-1)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}\frac{1}{F_{1-\alpha/2}(n_1-1, n_2-1)}\right\} = 1-\alpha.$$

于是得到 σ_1^2/σ_2^2 的置信区间为

$$\left(\frac{S_1^2}{S_2^2}\frac{1}{F_{\alpha/2}(n_1-1, n_2-1)}, \frac{S_1^2}{S_2^2}\frac{1}{F_{1-\alpha/2}(n_1-1, n_2-1)}\right).$$

例

为研究由机器 A 和机器 B 生产钢管的内径（以 mm 计），随机抽取机器 A 生产的管子 18 只，测得样本方差 $s_1^2 = 0.34$；随机抽取机器 B 生产的管子 13 只，测得样本方差 $s_2^2 = 0.29$。设两组样本相互独立，且由机器 A 和机器 B 生产的钢管内径分别服从正态分布 $N(\mu_1, \sigma_1^2)$ $N(\mu_2, \sigma_2^2)$，其中 μ_1, σ_i^2 $(i = 1, 2)$ 均未知。试求方差比 σ_1^2/σ_2^2 的置信水平为 0.90 的置信区间。

解：根据题意，有 $n_1 = 18$，$n_2 = 13$，$\alpha = 0.1$，查表得 $F_{0.05}(17, 12) = 2.59$，$F_{0.95}(17, 12) = \frac{1}{F_{0.05}(12,17)} = \frac{1}{2.38}$（$F$ 分布的性质，见书本 P. 145）。于是由前面分析得到所求置信区间为

$$\left(\frac{0.34}{0.29} \times \frac{1}{2.59}, \frac{0.34}{0.29} \times 2.38\right),$$

即 $(0.45, 2.79)$。由于该区间包含 1，在实际中我们就认为 σ_1^2, σ_2^2 两者没有显著差别（即比值“近似”为 1）。

主要内容

设有一组容量 $n>50$ 的大样本，它来自服从 $(0-1)$ 分布的总体 X，分布律（概率质量函数）为

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

其中 $p\in(0,1)$ 为未知参数。试求 p 的置信水平为 $1-\alpha$ 的置信区间。

已知 $(0-1)$ 分布的均值和方差分别为

$$\mu = p, \quad \sigma^2 = p(1-p).$$

设 $X_1,\ldots,X_n$ 为一组样本。由于样本容量 n 较大，由中心极限定理知

$$\frac{\sum_{i=1}^n X_i - np}{\sqrt{np(1-p)}} = \frac{n\bar{X} - np}{\sqrt{np(1-p)}} \overset{\text{近似}}{\sim} N(0,1).$$

于是有

$$P\left\{-z_{\alpha/2} < \frac{n\bar{X} - np}{\sqrt{np(1-p)}} < z_{\alpha/2}\right\} \approx 1-\alpha.$$

花括号中的不等式等价于

$$(n + z_{\alpha/2}^2)p^2 - (2n\bar{X} + z_{\alpha/2}^2)p + n\bar{X}^2 < 0. \tag{8}$$

把左边关于 p 的二次函数的系数记作

$$a = n + z_{\alpha/2}^2, \quad b = -(2n\bar{X} + z_{\alpha/2}^2), \quad c = n\bar{X}^2,$$

则由求根公式定义

$$p_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a},$$

即得到(8)的解为 $p_1 < p < p_2$.

因此，

$$P\{p_1 < p < p_2\} \approx 1-\alpha,$$

也就是说，p 的一个置信水平近似为 $1-\alpha$ 的置信区间为 (p_1, p_2)。

例

在一大批产品的 100 个样品中，检测得到一级品 60 个。求这批产品的一级品率 p 的置信水平为 0.95 的置信区间。

解：一级品率 p 是 $(0-1)$ 分布的参数。此处的样本空间为这一批的所有产品，随机变量

$$X(e) = \begin{cases} 1 & \text{如果}e\text{为一级品}, \\ 0 & \text{如果}e\text{不是一级品}. \end{cases}$$

本题 $n=100$，$\bar{x} = 60/100 = 0.6$，$1-\alpha = 0.95$，$\alpha/2 = 0.025$，$z_{\alpha/2} = 1.96$。按上面的分析来求 p 的置信区间，其中

$$a = n + z_{\alpha/2}^2 = 103.84, \quad b = -(2n\bar{X} + z_{\alpha/2}^2) = -123.84, \quad c = n\bar{X}^2 = 36.$$

于是得到 $p_1 = 0.50$，$p_2 = 0.69$。故 p 的一个近似置信区间为 $(0.50, 0.69)$。

Ex. 证明：不等式(8)左侧的二次函数有两个实根$p_{1,2}$ 且满足 $0 < p_1 < p_2 < 1$。因此 (p_1, p_2) 作为 $(0,1)$ 的子区间，确实能成为参数 $p \in (0,1)$ 的置信区间。

提示：在大样本假设下（即 n 充分大），根据大数定律，有 $\bar{X} \to \mu = p < 1$.

主要内容

在前面的讨论中，我们给出了未知参数 θ 的**双侧**置信区间 $(\underline{\theta},\overline{\theta})$，其中 $\underline{\theta},\overline{\theta}$ 为两个统计量。但在一些实际问题中，我们只关心**单侧**置信区间。

例如对于设备、原件的寿命，我们关心其平均值的下限；而考虑化学药品中的杂质含量时，我们关心其上限。

Definition (单侧置信区间)

对于给定值 $\alpha \in (0,1)$，若由样本确定的统计量 $\underline{\theta} = \underline{\theta}(X_1 \quad \ldots, X_n)$ 对于任意 $\theta \in \Theta$ 都满足

$$P\{\theta > \underline{\theta}\} \geq 1-\alpha,$$

则称随机区间 $(\underline{\theta},\infty)$ 是 θ 的置信水平为 $1-\alpha$ 的单侧置信区间，$\underline{\theta}$ 为单侧置信下限。
类似地，若统计量 $\overline{\theta} = \overline{\theta}(X_1 \quad \ldots, X_n)$ 对于任意 $\theta \in \Theta$ 都满足

$$P\{\theta < \overline{\theta}\} \geq 1-\alpha,$$

则也称随机区间 $(-\infty,\overline{\theta})$ 是 θ 的置信水平为 $1-\alpha$ 的单侧置信区间，$\overline{\theta}$ 为单侧置信上限。

例如，设 $X_1, \dots, X_n$ 是来自正态总体 X 的样本，其中总体均值 μ 和方差 σ^2 均未知。由

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$$

知

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

即

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

由此得到 μ 的置信水平为 $1-\alpha$ 的单侧置信区间

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

其中单侧置信下限

$$\underline{\mu} = \bar{X} - \frac{S}{\sqrt{n}} t_\alpha(n-1).$$

又由

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

知

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

即

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

由此得到 σ^2 的置信水平为 $1-\alpha$ 的单侧置信区间

$$\left(0,\ \frac{(n-1)S^2}{\chi^2_{1-\alpha}(n-1)}\right),$$

其中单侧置信上限为

$$\overline{\sigma^2} = \frac{(n-1)S^2}{\chi^2_{1-\alpha}(n-1)}.$$

例

从一批灯泡中随机取 5 只做寿命试验，测得寿命（以 h 计）为

$$1050 \quad 1100 \quad 1120 \quad 1250 \quad 1280$$

设灯泡寿命服从正态分布（评论：假设不合理，因为寿命非负）。求灯泡寿命均值的置信水平为 0.95 的单侧置信下限。

解：本题 $1-\alpha=0.95$，$n=5$，计算得 $\bar{x}=1160$，$s^2=9950$，查表得 $t_{0.05}(4)=2.1318$。把数据代入前面的分析结果，得到总体均值 μ 的单侧置信下限为

$$\underline{\mu}=\bar{x}-\frac{s}{\sqrt{n}}t_{0.05}(4)=1065.$$

The End

本章作业：TBA...