Parameter Estimation

December 12, 2024

引言

假设我们已知南京大学男性学生的身高服从正态分布 $\mathcal{N}(\mu, \sigma^2)$, 但不知道参数 μ 和 σ 具体的取值, 这时候我们可以利用抽样得到样本均值来推断总体的均值 μ.

这类已知总体分布形式, 但不知其具体参数, 用样本统计量来估计总体的参数的问题称为参数估计问题. 参数估计是统计推断的核心问题之一, 方法大体上有两类: 点估计与区间估计.

提纲：点估计

- 矩估计
- 极大似然估计
- 估计量的评价标准
  - 无偏性
  - 有效性：Rao-Crammer 不等式
  - 一致性

点估计

定义 0.84 设 $X_1, X_2, \ldots, X_n$ 是来自总体的一个样本, 用于估计未知参数 θ 的统计量 $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$ 称为 θ 的估计量, 或称为 θ 的 点估计.

Remarks: 点估计的本质就是用样本统计量直接作为总体参数的估计值

- 这里的参数是总体的属性, 而统计量是针对样本的计算.
- 在这里如何构造 $\hat{\theta}$ 没有明确的规定, 1900 年 K. 皮尔逊提出了一个替换原理, 给出了构造 $\hat{\theta}$ 的一种方法, 后来人们称此法为 矩估计法.
- 而 1922 年费希尔提出的最大似然法给出了另外一种构造 $\hat{\theta}$ 的方法, 称为 极大似然估计.

矩估计法

替换原理具体为:

- 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距).
  - 使用原点矩
    - 总体 X 的 k 阶原点矩: $a_k = \mathbb{E}[X^k]$
    - 样本的 k 阶原点矩: $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$
  - 使用中心矩
    - 总体 X 的 k 阶中心矩: $b_k = \mathbb{E}[(X - \mathbb{E}(X))^k]$
    - 样本的 k 阶中心矩: $B_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$
- 用样本矩的函数去替换相应的总体矩的函数.

矩估计法 -- 适用场景

根据这个替换原理, 在总体分布形式未知的情况下也可以对参数做出估计, 譬如:

- 用样本均值 $\bar{X}$ 估计总体均值 $\mathbb{E}(X)$
- 用样本方差 S^2 估计总体方差 $\mathbb{VAR}(X)$
  - 注意: 若没有特殊说明, 样本方差采用无偏方差
- 用事件 A 出现的频率估计事件 A 发生的概率.

矩估计法 -- 计算步骤

总体 X 的分布函数 F 包含 m 个未知参数 $\theta_1, \theta_2, \ldots, \theta_m$

- 计算总体 X 的 k 阶矩: $a_k = a_k(\theta_1, \theta_2, \ldots, \theta_m) = \mathbb{E}[X^k], k \in [m]$
  - a_k 一般为 $\theta_1, \theta_2, \ldots, \theta_m$ 的函数
- 计算样本的 k 阶矩: $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$
- 令样本矩等于总体矩:

$$A_k = a_k = a_k(\theta_1, \theta_2, \ldots, \theta_m), \quad k = [m]$$

  得到 m 个关于 $\theta_1, \theta_2, \ldots, \theta_m$ 的方程组

- 求解方程组得到估计量 $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m$

矩估计：例 0.154

例 0.154 设总体 X 的概率密度函数为

$$f(X)=\begin{cases}(\alpha+1)X^{\alpha}, & X\in(0,1)\\ 0, & \text{其它}\end{cases}$$

设 $X_1,X_2,\ldots,X_n$ 是来自总体的样本, 求参数 α 的矩估计.

解答：例 0.154

题目: 设总体 X 的概率密度函数为

$$f(X)=\begin{cases}(\alpha+1)X^{\alpha}, & X\in(0,1)\\ 0, & \text{其它}\end{cases}$$

设 $X_1,X_2,\dots,X_n$ 是来自总体的样本, 求参数 α 的矩估计.

解答:

- 首先计算总体 X 的 1 阶矩:

$$\mathbb{E}[X]=\int_{-\infty}^{+\infty}Xf(X)\,\mathrm{d}X=\int_0^1 X(\alpha+1)X^{\alpha}\,\mathrm{d}X=\frac{\alpha+1}{\alpha+2}.$$

  以及样本的 1 阶矩: $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$.

- 根据矩估计方法有

$$\mathbb{E}[X]=\frac{\alpha+1}{\alpha+2}=\bar{X}$$

  求解可得 $\alpha=(2\bar{X}-1)/(1-\bar{X})$.

矩估计：例 0.155

例 0.155 设 $X_1, X_2, \ldots, X_n$ 是来自总体的样本, 且总体 X 的概率密度函数为

$$f(X) = \begin{cases} \frac{1}{\theta} e^{-\frac{X-\mu}{\theta}}, & X \geq \mu \\ 0, & \text{其它} \end{cases}$$

其中 $\theta > 0$, 求参数 μ 和 θ 的矩估计.

解答：例 0.155

题目: 如上所述.

解答:

- 设随机变量 $Y = X - \mu$, 则 Y 服从参数为 $1/\theta$ 的指数分布, 有

$$\mathbb{E}[Y] = 0 \quad \text{和} \quad \sigma(Y) = \theta^2.$$

  由此可得 $\mathbb{E}[X] = \mu + \theta$ 和 $\sigma(X) = \theta^2$.

- 计算对应的样本矩

$$A_1 = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad B_2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

  求解方程组

$$\mu + \theta = A_1 \quad \text{和} \quad \theta^2 = B_2.$$

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

极大似然估计法 -- 例子

为了叙述极大似然估计的直观想法, 先看下面这个例子:

例 0.156 设有两个外形相同的箱子中各有 100 只球, 其中甲箱中有 99 个白球、1 个黑球, 乙箱中有 1 个白球、99 个黑球. 今随机抽取一箱并从中抽取一球, 结果取得白球, 问这个白球是从哪个箱子中取出?

极大似然估计法 – – 例子

为了叙述极大似然估计的直观想法, 先看下面这个例子:

例 0.157 设有两个外形相同的箱子中各有 100 只球, 其中甲箱中有 99 个白球、1 个黑球, 乙箱中有 1 个白球、99 个黑球. 今随机抽取一箱并从中抽取一球, 结果取得白球, 问这个白球是从哪个箱子中取出?

解答: A 表示事件 "从甲箱中取出白球", B 表示事件 "从乙箱中取出白球", 又

$$P(A) = 0.99 > P(B) = 0.01$$

因此, 按照可以推断白球 "最可能" 是从甲箱中取出的.

这个推断符合人们的经验事实, 这里的 "最可能" 就是 "极大似然" 之意, 这种想法常称为 "极大似然原理". 即, 已经得到了样本, 然后通过样本倒推, 找到能够使的该样本出现的最大概率的条件.

极大似然估计法

定义 0.85 设总体的概率函数为 $p(X;\theta), \theta\in\Theta$, 其中 θ 是一个未知参数或几个未知参数组成的参数向量, Θ 是参数空间. $X_1, X_2, \ldots, X_n$ 是来自总体的样本, 将样本的联合概率函数看成 θ 的函数, 用 $L(\theta, X_1, X_2, \ldots, X_n)$ 表示, 简记 $L(\theta)$,

$$L(\theta) = L(\theta, X_1, X_2, \ldots, X_n) = p(X_1;\theta)p(X_2;\theta)\ldots p(X_n;\theta)\ ,$$

$L(\theta)$ 称为样本的似然函数. 若某个统计量 $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$ 满足,

$$L(\hat{\theta}) = \max_{\theta\in\Theta}\ L(\theta)\ ,$$

则称 $\hat{\theta}$ 是 θ 的 极大似然估计, 简记为 MLE (Maximum Likelihood Estimation).

极大似然估计法 -- 计算步骤

求 $L(\theta) = p(X_1;\theta)p(X_2;\theta)\dots p(X_n;\theta)$ 的最大值可以通过下列步骤:

- 列出 $L(\theta) = p(X_1;\theta)p(X_2;\theta)\dots p(X_n;\theta)$
- 对等式两边取对数, 求关于 θ 求一阶偏导令其为零
- 求解方程组得到极大似然估计 $\hat{\theta}$.

极大似然估计法 -- 计算步骤

求 $L(\theta) = p(X_1;\theta)p(X_2;\theta)\ldots p(X_n;\theta)$ 的最大值可以通过下列步骤:

- 列出 $L(\theta) = p(X_1;\theta)p(X_2;\theta)\ldots p(X_n;\theta)$
- 对等式两边取对数, 求关于 θ 求一阶偏导令其为零
- 求解方程组得到极大似然估计 $\hat{\theta}$.

如何构造似然函数?

- 核心: 条件概率公式

极大似然估计：例 0.158

例 0.158 设 $X_1, X_2, \ldots, X_n$ 是取自总体 $X \sim B(1,p)$ 的样本, 求参数 p 的极大似然估计.

解答：例 0.158

题目: 设 $X_1, X_2, \ldots, X_n$ 是取自总体 $X \sim B(1,p)$ 的样本, 求参数 p 的极大似然估计.

解答:

- 首先计算似然函数

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

由而可得对数似然函数

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

求一阶偏导并令其为零, 可得

$$\frac{\partial \ln L(p)}{\partial p} = \frac{1}{p}\sum_{i=1}^{n} X_i - \frac{1}{1-p}\left(n - \sum_{i=1}^{n} X_i\right) = 0.$$

由此求解 $p = \sum_{i=1}^{n} X_i / n = \bar{X}$.

极大似然估计：例 0.159

例 0.159 设 $X_1, X_2, \ldots, X_n$ 是取自总体 $X \sim U(0,\theta)$ 的样本, 求参数 θ 的极大似然估计.

解答：例 0.159

题目: 设 $X_1, X_2, \ldots, X_n$ 是取自总体 $X \sim U(0,\theta)$ 的样本, 求参数 θ 的极大似然估计.

解答:

- 首先计算似然函数

$$L(\theta) = \frac{1}{\theta^n}\prod_{i=1}^{n}\mathbb{I}_{\{0<X_i\leq\theta\}} = \frac{1}{\theta^n}\mathbb{I}_{\{0<X_i\leq\theta\}}\ ,$$

  要使 $L(\theta)$ 最大, 首先是示性函数取值应该为 1, 其次是 $1/\theta^n$ 尽可能大, 由于 $1/\theta^n$ 是 θ 的单调递减函数, 所以 θ 的取值应尽可能小, 但示性函数为 1 决定了 θ 不能小于 $X_{(n)}$, 由此给出 θ 的极大似然估计为 $X_{(n)}$.

- 这个例子说明虽然求导函数是求极大似然估计最常用的方法, 但并不是所有场合求导都是有效的.

极大似然估计 -- 不可变性

极大似然估计有一个简单而有效的性质:

定理 0.74 如果 $\hat{\theta}$ 是参数 θ 的极大似然估计, 那么对于任一的函数 $g(\cdot)$, $g(\hat{\theta})$ 也是 $g(\theta)$ 的极大似然估计.

该性质称为极大似然估计的不变性, 从而使得一些复杂结构的参数的极大似然估计的计算变得容易了.

极大似然估计：例 0.160

例 0.160 设 $X_1, X_2, \ldots, X_n$ 是取自总体 $X \sim \mathcal{N}(\mu, \sigma^2)$ 的样本, 求参数 μ 和 $\sigma > 0$ 的极大似然估计.

解答：例 0.160

解答:

- 根据正态分布的密度函数, 可知似然函数

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

其对数似然函数为 $\ln L(\mu,\sigma)=-n\ln(2\pi)^{1/2}-n\ln\sigma-\sum_{i=1}^{n}(X_i-\mu)^2/2\sigma^2$.

- 对参数 μ 求导计算可得

$$\frac{\partial\ln L(\mu,\sigma)}{\partial\mu}=\sum_{i=1}^{n}(X_i-\mu)=0\Rightarrow\mu=\frac{1}{n}\sum_{i=1}^{n}X_i=\bar{X},$$

对参数 σ 求导计算, 可得

$$\frac{\partial\ln L(\mu,\sigma)}{\partial\sigma}=-\frac{n}{\sigma}+\frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i-\mu)^2=0\Rightarrow\sigma=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2}.$$

根据极大似然估计的不变性, 可知方差 σ^2 的极大似然估计为

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

极大似然估计：例 0.161

例 0.161 设 $X_1, X_2, \dots, X_n$ 是来自总体的样本, 且总体 X 的概率密度函数为

$$f(X) = \begin{cases} \theta e^{-(X-\mu)\theta}, & X \geq \mu \\ 0, & \text{其它} \end{cases}$$

求参数 μ 和 θ 的极大似然估计.

解答：例 0.161

题目: 设 $X_1, X_2, \ldots, X_n$ 是来自总体的样本, 且总体 X 的概率密度函数为

$$f(X) = \begin{cases} \theta e^{-(X-\mu)\theta}, & X \geq \mu \\ 0, & \text{其它} \end{cases}$$

求参数 μ 和 θ 的极大似然估计.

解答:

- 列出似然函数

$$L(\theta, \mu) = \begin{cases} \theta^n e^{-\theta \sum_{i=1}^n (X_i - \mu)}, & X_i \geq \mu \\ 0, & \text{其它} \end{cases}$$

  其对数似然函数为

$$\ln L(\theta, \mu) = n \ln \theta - \theta \sum_{i=1}^n (X_i - \mu).$$

- 对参数 θ 求导计算可得

$$\frac{\partial \ln L(\theta,\mu)}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^{n}(X_i - \mu) = 0 \Rightarrow \theta = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)},$$

对参数 μ 求导计算可得

$$\frac{\partial \ln L(\theta,\mu)}{\partial \mu} = n\theta = 0 \Rightarrow \theta = 0.$$

此时无法求解 μ 和 θ 的极大似然估计.

- 回顾似然函数

$$L(\theta,\mu) = \begin{cases} \theta^n e^{-\theta\sum_{i=1}^{n}(X_i-\mu)}, & X_i \geq \mu \\ 0, & \text{其它} \end{cases}$$

可以发现 μ 越大似然函数 $L(\theta,\mu)$ 越大, 但须满足 $X_i \geq \mu (i \in [n])$. 由此可得极大似然估计为

$$\hat{\mu} = X_{(1)}, \qquad \hat{\theta} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}(X_i - X_{(1)})}.$$

极大似然估计：例 0.162

例 0.162 设总体 X 的概率密度函数为

$$f(X)=\begin{cases}(\alpha+1)X^{\alpha}, & X\in(0,1)\\ 0, & \text{其它}\end{cases}$$

设 $X_1,X_2,\ldots,X_n$ 是来自总体的样本, 求参数 α 的矩估计.

解答：例 0.162

题目: 如上所述.

解答:

- 列出似然函数

$$L(\alpha) = (\alpha+1)^n \prod_{i=1}^{n} X_i^\alpha = (\alpha+1)^n (X_1 X_2 \ldots X_n)^\alpha,$$

以及其对数似然函数为 $\ln L(\alpha) = n\ln(\alpha+1) + \alpha\ln(X_1X_2\ldots X_n)$. 求导并令导数为零有

$$\frac{\partial \ln L(\alpha)}{\partial \alpha} = \frac{n}{\alpha+1} + \ln(X_1X_2\ldots X_n) = 0,$$

求解可得

$$\alpha = \frac{-n}{\sum_{i=1}^{n}\ln(X_i)} - 1$$

对比例 0.154, 可以看到同一密度函数的矩估计和极大似然估计结果可能不同.

估计量的评价标准

不同的估计方法可能得到不同的估计值.

自然地, 我们希望知道采用哪一种估计量更好, 或更好的标准是什么呢? 统计学上, 给出了无偏性、有效性、一致性等评价标准.

- 无偏性: $\hat{\theta}$ 与参数真值 θ 之间的偏差的平均值为 0
- 有效性: $\hat{\theta}$ 围绕参数真值 θ 的方差越小越好
- 一致性: 随着样本量的不断增大, $\hat{\theta}$ 能够有效逼近参数真值 θ

无偏性

定义 0.86 设 $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$ 是 θ 的一个估计, θ 的参数空间为 Θ, 若对任意的 $\theta \in \Theta$, 有

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

则称 $\hat{\theta}$ 是 θ 的 无偏估计, 否则称为 有偏估计.

Remarks:

- (原点矩) 样本 k 阶原点矩为总体 k 阶原点矩的无偏估计
- (中心矩) 设 $X_1, X_2, \ldots, X_n$ 来自总体 X 的样本, 期望 μ, 方差 σ^2, 则:
  - $S_0^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$ 是 σ^2 的有偏估计
  - $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ 是 σ^2 的无偏估计
- 若 $\hat{\theta}$ 是 θ 的一个无偏估计, $g(\hat{\theta})$ 不一定也是 $g(\theta)$ 的无偏估计.

无偏估计：例 0.163

例 0.163 设 $X_1, X_2, \ldots, X_n$ 来自总体 X 的样本, 且总体 X 的概率密度函数为

$$f(X) = \begin{cases} \frac{1}{\theta} e^{-\frac{X}{\theta}}, & X \geq 0 \\ 0, & X < 0 \end{cases}$$

证明: 统计量

$$\bar{X} = \sum_{i=1}^{n} X_i / n \quad \text{和} \quad n \cdot \min\{X_1, X_2, \ldots, X_n\}$$

均是 θ 的无偏估计.

解答：例 0.163

题目: 如上所述.

解答:

- 根据期望和指数分布的性质, 有

$$\mathbb{E}[\bar{X}] = \mathbb{E}[X] = \theta,$$

  由此可知, $\bar{X}$ 是 θ 的无偏估计 (原点矩) .
- 设随机变量 $Z = \min\{X_1, X_2, \ldots, X_n\}$, 则有

$$\begin{aligned} F_Z(z) &= P(Z \leq z) = 1 - P(Z > z) \\ &= 1 - P(X_1 > z)P(X_X > z)\ldots P(X_n > z) \\ &= 1 - \prod_{i=1}^{n}(1 - P(X_i \leq z)) = \begin{cases} 0, & z < 0 \\ 1 - e^{-nz/\theta}, & z \geq 0 \end{cases} \end{aligned}$$

于是当 $z \geq 0$ 时, 有

$$P(Z > z) = 1 - F_Z(z) = e^{-nz/\theta}.$$

根据期望的性质, 有

$$\mathbb{E}[Z] = \int_0^{+\infty} e^{-nz/\theta}\,\mathrm{d}z = \frac{\theta}{n},$$

于是有 $\theta = \mathbb{E}[nZ]$.

有效性

例子 0.163 说明: 参数可能存在多个无偏估计.

- 若 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 都是 θ 的无偏估计, 如何在多个无偏估计中进行选择?
- 直观的想法是, $\hat{\theta}$ 围绕参数真值 θ 的方差越小越好, 即有效性.

定义 0.87 设 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 分别是 θ 的两个无偏估计, 如果对任意的 $\theta \in \Theta$ 都有

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

且至少有一个 $\theta \in \Theta$ 使得上述不等式严格成立, 则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 有效.

Remarks: 有效性是针对无偏估计而言的, 因此判断有效性之前必须先确认估计量的无偏性.

有效性：例 0.164

例 0.164 设 $X_1, X_2, \ldots, X_n$ 来自总体 X 的样本, 且总体 X 的概率密度函数为

$$f(X) = \begin{cases} \frac{1}{\theta} e^{-\frac{X}{\theta}}, & X \geq 0 \\ 0, & X < 0 \end{cases}$$

令 $Z = \min\{X_1, X_2, \ldots, X_n\}$. 证明: 当 $n > 1$ 时, $\bar{X} = \sum_{i=1}^{n} X_i / n$ 比 nZ 更有效.

解答：例 0.164

题目: 如上所述.

解答:

- 根据样本的独立性有
$$\sigma(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma(X_i) = \frac{\theta^2}{n}.$$
又根据例0.163可知随机变量 Z 的密度函数为
$$f(z) = \begin{cases} 0, & z < 0 \\ \frac{n}{\theta}e^{-\frac{nz}{\theta}}, & z \geq 0 \end{cases}$$
从而得到
$$\sigma(nZ) = n^2\sigma(Z) = n^2\frac{\theta^2}{n^2} = \theta^2,$$
因此当 $n \geq 1$ 时有 $\sigma(\bar{X}) \leq \sigma(nZ)$ 成立, 故估计量 $\bar{X}$ 比 nZ 更有效.

有效性：例 0.165

例 0.165 设 $X_1, X_2, \ldots, X_n$ 是来自总体 X 的样本, 且 $\mathbb{E}(X) = \mu$ 及 $\mathbb{VAR}(X) = \sigma^2$. 设常数 $c_1, c_2, \ldots, c_n \geq 0$, 满足 $\sum_{i=1}^n c_i = 1, c_i \neq \frac{1}{n}$. 求证: $\bar{X}$ 比 $\sum_{i=1}^n c_i X_i$ 有效.

解答：例 0.165

题目: 设 $X_1, X_2, \ldots, X_n$ 是来自总体 X 的样本, 且 $\mathbb{E}(X) = \mu$ 及 $\mathbb{VAR}(X) = \sigma^2$. 设常数 $c_1, c_2, \ldots, c_n \geq 0$, 满足 $\sum_{i=1}^n c_i = 1, c_i \neq \frac{1}{n}$. 求证: $\bar{X}$ 比 $\sum_{i=1}^n c_i X_i$ 有效.

解答:

- 根据样本的独立同分布的性质, 有

$$\mathbb{E}(\bar{X}) = \mu \quad \text{和} \quad \mathbb{VAR}(\bar{X}) = \frac{\sigma^2}{n}.$$

根据期望的性质, 有 $\mathbb{E}[\sum_{i=1}^n c_i X_i] = \mu$, 进一步有

$$\mathbb{VAR}\left(\sum_{i=1}^n c_i X_i\right) = \sum_{i=1}^n c_i^2 \mathbb{VAR}(X_i) = \sigma^2 \sum_{i=1}^n c_i^2 \geq \frac{\sigma^2}{n}.$$

这里利用不等式 $\sum_{i=1}^n c_i^2/n \geq (\sum_{i=1}^n c_i/n)^2$, 所以有 $\mathbb{VAR}(\sum_{i=1}^n c_i X_i) \geq \mathbb{VAR}(\bar{X})$.

Rao-Crammer 不等式

有效性希望 $\hat{\theta}$ 围绕参数真值 θ 的方差越小越好, 那么这个方差能小到什么程度? 有无下界? 若有的话, 如何去求? Rao-Crammer 不等式回答了这些问题.

定理 0.75 随机变量 X 的概率密度为 $f(X;\theta)$ 或分布函数为 $F(X;\theta)$, 令

$$\mathbb{VAR}_0(\theta)=\frac{1}{n\mathbb{E}\left[\left(\frac{\partial \ln f(X;\theta)}{\partial\theta}\right)^2\right]} \quad \text{或} \quad \mathbb{VAR}_0(\theta)=\frac{1}{n\mathbb{E}\left[\left(\frac{\partial \ln F(X;\theta)}{\partial\theta}\right)^2\right]}$$

对任意的无偏估计量 $\hat{\theta}$, 有

$$\mathbb{VAR}(\hat{\theta})\geq\mathbb{VAR}_0(\theta)\ ,$$

称 $\mathbb{VAR}_0(\theta)$ 为估计量 $\hat{\theta}$ 方差的下界. 当 $\mathbb{VAR}(\hat{\theta})=\mathbb{VAR}_0(\theta)$ 时, 称 $\hat{\theta}$ 为达到方差下界的无偏估计量, 此时 $\hat{\theta}$ 为最有效估计量, 简称有效估计量.

有效性：例 0.166

例 0.166 设 $X_1, X_2, \ldots, X_n$ 是来自总体 X 的样本, 且总体 X 的概率密度函数为

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

证明: θ 的极大似然估计为有效估计量.

解答：例 0.166

题目: 如上所述.

解答:

- 根据定理 0.75, 首先计算 $\sigma_0(\theta)$. 又
$$\ln f(X;\theta) = -\ln\theta - \frac{X}{\theta}, \qquad \frac{\partial \ln f(X;\theta)}{\partial\theta} = -\frac{1}{\theta} + \frac{X}{\theta^2}$$
所以
$$\mathbb{VAR}_0(\theta) = \frac{1}{n\mathbb{E}\left[\left(\frac{\partial \ln f(X;\theta)}{\partial\theta}\right)^2\right]} = \frac{1}{n\mathbb{E}\left[\left(-\frac{1}{\theta} + \frac{X}{\theta^2}\right)^2\right]} = \frac{1}{\frac{n}{\theta^4}\mathbb{E}[(X-\mathbb{E}[X])^2]} = \frac{\theta^2}{n}$$
- 计算对数似然函数, 有
$$\ln L(\theta) = -n\ln\theta - \frac{1}{\theta}\sum_{i=1}^{n} X_i \Rightarrow \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i\ ,$$
进一步得到极大似然估计 $\hat{\theta}$ 的方差 $\mathbb{VAR}(\hat{\theta}) = \frac{\theta^2}{n}$, 因此 θ 的极大似然估计为有效估计量.

一致性

定义 0.88 设 $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$ 是 θ 一个估计量. 当 $n \to \infty$ 时, 有 $\hat{\theta} \xrightarrow{P} \theta$ 成立, 即对任意 $\epsilon > 0$, 有

$$\lim_{n\to 0} P\left[|\hat{\theta} - \theta| > \epsilon\right] = 0\,,$$

则称 $\hat{\theta}$ 为 θ 的 一致估计量.

Remarks: 一致性被认为是对估计的一个最基本要求.

- 如果一个估计量, 在样本不断增多时都不能有效的靠近被估参数的真实值, 那么这个估计是很值得怀疑的.
- 通常, 不满足一致性的估计都不予考虑.

一致性

在判断或计算参数的一致估计量时, 下述两个定理是很有用的.

定理 0.76 (一致性的充分条件) 设 $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$ 是 θ 的一个估计量, 若满足以下两个条件:

$$\lim_{n\to\infty} \mathbb{E}[\hat{\theta}_n] = \theta\ , \qquad \lim_{n\to\infty} \mathbb{VAR}\left[\hat{\theta}_n\right] = 0$$

则 $\hat{\theta}$ 为 θ 的一致估计量.

定理 0.77 (一致性的函数不变性) 设 $\hat{\theta}_{n1}, \hat{\theta}_{n2}, \ldots, \hat{\theta}_{nk}$ 分别是 $\theta_1, \theta_2, \ldots, \theta_k$ 的一致性估计, $G = g(\theta_1, \theta_2, \ldots, \theta_k)$ 是 $\theta_1, \theta_2, \ldots, \theta_k$ 的连续函数, 则

$$\hat{G} = g\left(\hat{\theta}_{n1}, \hat{\theta}_{n2}, \ldots, \hat{\theta}_{nk}\right)$$

是 G 的一致性估计.

一致性：例 0.167

例 0.167 设 $X_1, X_2, \ldots, X_n$ 是来自总体 X 的样本, 且总体 X 的概率密度函数为

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

证明: 样本均值 $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 是 θ 的一致估计量.

解答：例 0.167

题目: 设 $X_1, X_2, \ldots, X_n$ 是来自总体 X 的样本, 且总体 X 的概率密度函数为

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

证明: 样本均值 $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 是 θ 的一致估计量.

解答:

- 根据定理 0.76, 有

$$\lim_{n\to\infty} \mathbb{E}[\hat{\theta}] = \theta, \qquad \lim_{n\to\infty} \mathbb{VAR}[\hat{\theta}] = \lim_{n\to\infty} \frac{\theta^2}{n} = 0.$$

证毕.

一致性：例 0.168

例 0.168 设 $X_1, X_2, \ldots, X_n$ 是取自总体 $X \sim U(0, \theta)$ 的样本. 证明: 参数 θ 的极大似然估计是一致估计量.

解答：例 0.168

- 根据前面的例题, 可知 θ 的极大似然估计是 $\hat{\theta} = x_{(n)}$. 设随机变量 $Z = x_{(n)}$, 则 Z 的分布函数为

$$F_Z(z) = P(Z \le z) = P(x_{(n)} \le z) = \prod_{i=1}^{n} P(x_i \le z) = \begin{cases} 1, & z > \theta \\ (\frac{z}{\theta})^n, & z \in [0, \theta] \\ 0, & z < 0 \end{cases}$$

由此可得当 $z \in [0, \theta]$ 时随机变量 Z 的密度函数为 $f_Z(z) = nz^{n-1}/\theta^n$.

- 进一步有

$$\mathbb{E}[\hat{\theta}] = \int_0^\theta \frac{nz^n}{\theta^n}\,\mathrm{d}z = \frac{n}{n+1}\theta \Rightarrow \lim_{n\to\infty} \mathbb{E}[\hat{\theta}] = \theta$$

又 $\mathbb{E}[Z^2] = \int_0^\theta \frac{nz^{n+1}}{\theta^n}\,\mathrm{d}z = \frac{n}{n+2}\theta^2$, 因此有

$$\mathbb{VAR}[\hat{\theta}] = \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2 = \frac{n}{n+2}\theta^2 - (\frac{n\theta}{n+1})^2 = \frac{n}{(n+1)^2(n+2)}\theta^2 \Rightarrow \lim_{n\to\infty} \mathbb{VAR}[\hat{\theta}] = 0$$

由此, $\hat{\theta}$ 是 θ 的有偏、但一致估计量.

Interval Estimation

December 12, 2024

区间估计

点估计用样本统计统计量直接作为总体参数的估计值, 简洁却存在以偏概全的局限性. 但是推断一个参数落在某个区间比推断这个参数具体的数值要简单. 因此, 我们可以推断总体的参数的是在某个区间范围内的. 比如通过 1 万人的样本预估出中国总体人口的平均身高是 165-175cm, 这就是区间估计.

区间的范围很大, 你可以预估身高是 165-175cm 之间, 也可以预估是 160-180cm 之间, 也可以是其他. 但你会看到, 前者相比后者预测准确的概率更低, 因为其预测的区间范围太窄, 而后者预测的区间范围更宽. 所以, 在进行区间估计的时候, 你会发现每一个预估的区间都对应一个预估的准确度. 前者被称为置信区间, 后者被称为置信度.

提纲：区间估计

- 置信区间与置信度
- 枢轴变量法
  - 单个正态总体参数：σ 已知时，μ 的置信区间
  - 单个正态总体参数：σ 未知时，μ 的置信区间
  - 单个正态总体参数：σ^2 的置信区间
  - 两个正态总体参数
  - 非正态分布的区间估计
- 单侧置信区间

置信区间与置信度

定义 0.89 设 $X_1, X_2, \ldots, X_n$ 是来自总体 X 的样本, 总体 X 的分布函数含未知参数 θ, 求统计量 $\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \ldots, X_n)$ 和 $\hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \ldots, X_n)$ 使得

$$\Pr[\hat{\theta}_1 < \theta < \hat{\theta}_2] \geq 1 - \alpha$$

成立, 则称 $1 - \alpha$ 为置信度, $[\hat{\theta}_1, \hat{\theta}_2]$ 为 θ 的置信度为 $1 - \alpha$ 的置信区间.

Remarks:

- 置信区间 $\hat{\theta}_2 - \hat{\theta}_1$ 反映了估计精度: 长度越小, 精度越大.
- 置信度 α 反映了估计的可靠度: α 越小, 可靠度越高.
- 给定置信度 α, 置信区间 $[\hat{\theta}_1, \hat{\theta}_2]$ 的选取并不唯一确定,. 通常选长度最小的一个区间.

枢轴变量法

构造未知参数 θ 的置信区间最常用的方法是枢轴变量法, 其步骤可以概括如下:

- 设法构造一个样本统计量 $W = W(X_1, X_2, \ldots, X_n; \theta)$, 使得 W 的分布包含待估参数 θ, 但不依赖其它参数, 且函数 W 的分布已知, W 称为枢轴变量.
- 给定置信度 $1-\alpha$, 根据 W 的分布找出临界值 a 和 b, 使得下式成立

$$\Pr[a < W < b] = 1-\alpha\,.$$

- 根据 $a < W < b$, 解出 $\hat{\theta}_1 < \theta < \hat{\theta}_2$, 则 $[\hat{\theta}_1, \hat{\theta}_2]$ 为 θ 的置信度为 $1-\alpha$ 的置信区间.

单个正态总体参数 -- σ 已知时 μ 的置信区间

按照枢轴变量法, 有:

- 由于 μ 的点估计为 $\bar{X}$, 且 $\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$, 因此枢轴变量可选为 $W = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$.
- 给定置信度 $1-\alpha$, 根据 W 的分布找出临界值 a 和 b, 使得

$$\Pr[a < W < b] = 1 - \alpha\,.$$

- 根据正态分布的性质、对称性和上分位点可知

$$\Pr[W \geq \mu_{1-\alpha/2}] = 1 - \alpha/2 \quad 和 \quad \Pr[W \leq \mu_{\alpha/2}] = 1 - \alpha/2\,.$$

求解可得 $a = \mu_{1-\alpha/2} = -\mu_{\alpha/2}$ 和 $b = \mu_{\alpha/2}$. 又因为 $W = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$, 于是有

$$\Pr[-\mu_{\alpha/2} < W < \mu_{\alpha/2}] = \Pr\left[\bar{X} - \frac{\sigma}{\sqrt{n}}\mu_{\alpha/2} < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}}\mu_{\alpha/2}\right] = 1-\alpha\,.$$

区间估计：例 0.169

例 0.169 用天平称量某物体的质量 9 次, 得平均值 $\bar{x}$ =15.4(g), 已知天平称量结果为正态分布, 其标准差为 0.1(g). 试求该物体质量的 0.95 置信区间.

解答：例 0.169

题目: 用天平称量某物体的质量 9 次, 得平均值 $\bar{x}$ =15.4(g), 已知天平称量结果为正态分布, 其标准差为 0.1(g). 试求该物体质量的 0.95 置信区间.

解答:

- 根据题意知 $1-\alpha=0.95, \alpha=0.05$, 查标准正态分布函数表知 $\mu_{\alpha/2}=\mu_{0.975}=1.96$.
- 进一步有该物体质量的 0.95 置信区间为

$$\bar{X} \pm \frac{\sigma}{\sqrt{n}}\mu_{\alpha/2}=15.4 \pm 1.96 \times 0.1/\sqrt{9}=15.4 \pm 0.0653.$$

从而该物体质量的 0.95 置信区间为 [15.3347, 15.4653].

区间估计：例 0.170

例 0.170 设总体为正态分布 $\mathcal{N}(\mu, 1)$, 为得到 μ 的置信度为 0.95 的置信区间且区间长度不超过 1.2, 求样本容量应为多大?

解答：例 0.170

题目: 设总体为正态分布 $\mathcal{N}(\mu,1)$, 为得到 μ 的置信度为 0.95 的置信区间且区间长度不超过 1.2, 求样本容量应为多大?

解答:

- 根据题意知 $1-\alpha=0.95, \alpha=0.05$, 查标准正态分布函数表知 $\mu_{\alpha/2}=\mu_{0.975}=1.96$.
- 又因为 μ 的置信度为 0.95 的置信区间为

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

其区间长度为 $2\mu_{\alpha/2}/\sqrt{n}$, 它仅依赖于样本容量 n 且与样本取值无关. 又

$$2\mu_{\alpha/2}/\sqrt{n}\leq 1.2 \Rightarrow n\geq (2/1.2)^2\mu_{\alpha/2}^2=10.67\approx 11.$$

即样本容量为 11 时, 使得 μ 的置信度为 0.95 的置信区间长度不超过 1.2.

单个正态总体参数 -- σ 未知时 μ 的置信区间

按照枢轴变量法, 有:

- σ^2 可用样本方差 $S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)$ 估计, 枢轴变量选为 $W = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$.
- 给定置信度 $1-\alpha$, 根据 W 的分布找出临界值 a 和 b, 使得

$$\Pr[a < W < b] = 1-\alpha \Rightarrow b = t_{\alpha/2}(n-1)\,, \quad a = -t_{\alpha/2}(n-1)\,.$$

- 整理可得

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

区间估计：例 0.171

例 0.171 假设轮胎的寿命服从正态分布. 为估计某种轮胎的平均寿命, 现随机地抽取 12 只轮胎试用, 测得它们的寿命 (单位: 万千米) 如下:

$$\begin{array}{cccccc} 4.68 & 4.85 & 4.32 & 4.85 & 4.61 & 5.02 \\ 5.20 & 4.60 & 4.58 & 4.72 & 4.38 & 4.7 \end{array}$$

试求平均寿命的 0.95 置信区间.

解答：例 0.171

题目: 假设轮胎的寿命服从正态分布. 为估计某种轮胎的平均寿命, 现随机地抽取 12 只轮胎试用, 测得它们的寿命 (单位: 万千米) 如下:

4.68	4.85	4.32	4.85	4.61	5.02
5.20	4.60	4.58	4.72	4.38	4.7

试求平均寿命的 0.95 置信区间.

解答:

- 此处正态总体标准差未知, 可使用 t 分布求均值的置信区间. 本例中经计算有

$$\bar{x} = 4.709, \qquad s^2 = 0.0615$$

取 $\alpha = 0.05$, 查表知 $t_{0.975}(11) = 2.2010$, 于是平均寿命的 0.95 置信区间为

$$4.7092 \pm 2.2010 \cdot \sqrt{0.0615}/\sqrt{12} = [4.5516, 4.8668]$$

单个正态总体参数 -- σ^2 的置信区间

虽然可以就 μ 是否已知分两种情况讨论 σ^2 的置信区间, 但在实际中 σ^2 未知 μ 已知的情形是极为罕见的, 所以我们只在 μ 未知的情况下讨论 σ^2 的置信区间. 按照枢轴变量法, 有:

- σ^2 可用样本方差 $S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2/(n-1)$ 估计, 枢轴变量选为 $W = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.
- 给定置信度 $1-\alpha$, 设临界值 a 和 b, 使得 $P[a < W < b] = 1-\alpha$. 又根据 χ^2 分布的不对称性, 采用概率对称的区间

$$P[W \le a] = P[W \ge b] = \frac{\alpha}{2} \Rightarrow b = \chi^2_{\alpha/2}(n-1)\,, \quad a = \chi^2_{1-\alpha/2}(n-1)\,.$$

- 整理可得

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

区间估计：例 0.172

例 0.172 某厂生产的零件质量服从正态分布 $\mathcal{N}(\mu, \sigma^2)$, 现随机地抽取 9 个零件, 测得其质量 (单位: g) 如下:

$$45.3 \quad 45.4 \quad 45.1 \quad 45.3 \quad 45.5 \quad 45.7 \quad 45.4 \quad 45.3 \quad 45.6$$

试求总体标准差 σ 的 0.95 置信区间.

解答：例 0.172

题目: 某厂生产的零件质量服从正态分布 $\mathcal{N}(\mu, \sigma^2)$, 现随机地抽取 9 个零件, 测得其质量 (单位: g) 如下:

45.3　45.4　45.1　45.3　45.5　45.7　45.4　45.3　45.6

试求总体标准差 σ 的 0.95 置信区间.

解答:

- 本例中经计算有 $(n-1)S^2 = 8\times 0.0325 = 0.26$, 又 $\alpha = 0.05$, 查表知 $\chi^2_{0.025}(8) = 2.1797$, $\chi^2_{0.975}(8) = 17.5345$, 于是总体标准差 σ^2 的 0.95 置信区间为

$$\left[\frac{0.26}{17.5345}, \frac{0.26}{2.1797}\right] = [0.0148, 0.1193]$$

从而 σ 的 0.95 置信区间为 $[0.1218, 0.3454]$.

两个正态总体参数的置信区间

若 $X_1, X_2, \ldots, X_n$ 是来自 $\mathcal{N}(\mu_1, \sigma_1^2)$ 和 $y_1, y_2, \ldots, y_m$ 是来自 $\mathcal{N}(\mu_2, \sigma_2^2)$ 的样本, 且两个样本相互独立, 令

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} X_i\,, \quad \bar{y} = \frac{1}{m}\sum_{i=1}^{m} y_i\,, \quad S_1^2 = \sum_{i=1}^{n}\frac{(X_i-\bar{x})^2}{(n-1)}, \quad S_2^2 = \sum_{i=1}^{m}\frac{(y_i-\bar{y})^2}{(m-1)}\,.$$

下面讨论两个均值之差 $\mu_1 - \mu_2$ 和两个方差之比 σ_1^2/σ_2^2 的置信度为 $1-\alpha$ 的区间估计.

两个正态总体参数的置信区间

- 当 σ_1^2 和 σ_2^2 已知时, 有

$$\bar{x} - \bar{y} \sim \mathcal{N}\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right) ,$$

取枢轴量为

$$W = \frac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim \mathcal{N}(0, 1) ,$$

求解置信区间, 有

$$\Pr\left[\bar{X} - \bar{Y} - \mu_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} < \mu_1 - \mu_2 < \bar{X} - \bar{Y} + \mu_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\right] = 1-\alpha \,.$$

两个正态总体参数的置信区间

- 当 $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 未知时, 有

$$\bar{x} - \bar{y} \sim \mathcal{N}\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right), \quad \frac{(n-1)S_1^2 + (m-1)S_2^2}{\sigma^2} \sim \chi^2(m+n-2),$$

故可以构造如下服从 t 分布的枢轴量

$$W = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_W\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t(n+m-2),$$

其中 $S_W^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}$. 于是置信区间为

$$\Pr\left[-t_{\alpha/2}(n+m-2) < \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_W\sqrt{\frac{1}{n} + \frac{1}{m}}} < t_{\alpha/2}(n+m-2)\right] = 1-\alpha.$$

两个正态总体参数的置信区间

- 求 σ_1^2/σ_2^2 的置信度为 $1-\alpha$ 的区间估计, 由于 $\frac{(n-1)S_1^2}{\sigma_1^2} \sim \chi^2(n-1)$, $\frac{(m-1)S_2^2}{\sigma_2^2} \sim \chi^2(m-1)$, 故可构造服从 F 分布的枢轴量

$$W = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n-1, m-1) ,$$

根据 F 分布的不对称性, 采用概率对称的区间

$$\Pr[W \leq a] = \Pr[W \geq b] = \alpha/2$$

$$\Rightarrow b = F_{\frac{\alpha}{2}}(n-1, m-1), \quad a = F_{1-\alpha/2}(n-1, m-1) .$$

由此可得置信区间

$$\Pr\left[\frac{S_1^2/S_2^2}{F_{\alpha/2}(n-1, m-1)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2/S_2^2}{F_{1-\alpha/2}(n-1, m-1)}\right] = 1-\alpha .$$

区间估计：例 0.173

例 0.173 若从总体 $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ 和总体 $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ 中分别抽取容量为 $n_1 = 10, n_2 = 15$ 的样本, 经计算后有 $\bar{X} = 82, S_1^2 = 56.5, \bar{Y} = 76, S_2^2 = 52.4.$

- 已知 $\sigma_1^2 = 64, \sigma_2^2 = 49$, 求 $\mu_1 - \mu_2$ 的 0.95 置信区间.
- 当 $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 未知时, 求 $\mu_1 - \mu_2$ 的 0.95 置信区间.
- 求 σ_1^2/σ_2^2 的 0.95 置信区间.

解答：例 0.173

解答:

- σ_1^2 和 σ_2^2 已知时的置信区间为

$$\Pr\left[\bar{X}-\bar{Y}-\mu_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}<\mu_1-\mu_2<\bar{X}-\bar{Y}+\mu_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}\right]=1-\alpha.$$

查标准正态分布函数表知 $\mu_{0.975}=1.96$, 因此 $\mu_1-\mu_2$ 的 **0.95** 置信区间为

$$\left[6-1.96\sqrt{\frac{64}{10}+\frac{49}{15}},6+1.96\sqrt{\frac{64}{10}+\frac{49}{15}}\right]=[-0.0939,12.0939]$$

- $\sigma_1^2=\sigma_2^2$ 时, 置信区间为

$$\Pr\left[-t_{\alpha/2}(n+m-2)<\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_W\sqrt{\frac{1}{n}+\frac{1}{m}}}<t_{\alpha/2}(n+m-2)\right]=1-\alpha.$$

其中 $S_W=\frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2}=\frac{9\times56.5+14\times52.4}{23}=54.0043$, 而 $t_{0.975}(23)=2.0687$, 因此

$\mu_1-\mu_2$ 的 0.95 置信区间为

$$\left[82-76-2.0687\sqrt{54.0043}\sqrt{\frac{10+15}{10\times 15}}, 82-76+2.0687\sqrt{54.0043}\sqrt{\frac{10+15}{10\times 15}}\right]$$
$$=[-0.2063, 12.2063].$$

- σ_1^2/σ_2^2 的置信度为 $1-\alpha$ 的置信区间为

$$\Pr\left[\frac{S_1^2/S_2^2}{F_{\alpha/2}(n-1,m-1)}<\frac{\sigma_1^2}{\sigma_2^2}<\frac{S_1^2/S_2^2}{F_{1-\alpha/2}(n-1,m-1)}\right]=1-\alpha.$$

查表得 $F_{0.975}(9,14)=3.21$, $F_{0.025}(9,14)=1/F_{0.975}(14,9)=1/3.8$, 因此 σ_1^2/σ_2^2 的 0.95 置信区间为

$$\left[\frac{56.5}{52.4}\cdot\frac{1}{3.21}, \frac{56.5}{52.4}\cdot 3.8\right]=[0.3359, 4.0973]$$

非正态分布的区间估计

若总体 X 的分布未知或非正态分布, 我们可以利用集中不等式和中心极限定理给出总体期望 $\mu = \mathbb{E}[X]$ 的区间估计.

- 若 $X \in [a,b]$, 设 $\bar{X} = \sum_{i=1}^{n} X_i/n$, 根据集中不等式有

  $$\Pr\left[|\mu - \bar{X}| \geq \epsilon\right] \leq 2\exp(-2n\epsilon^2/(b-a)^2) \,.$$

  令

  $$\alpha = 2\exp(-2n\epsilon^2/(b-a)^2) \,,$$

  求解

  $$\epsilon = \sqrt{(b-a)^2 \ln(2/\alpha)/n} \,,$$

  于是有

  $$\Pr\left[\bar{X} - \sqrt{(b-a)^2 \ln(2/\alpha)/n} < \mu < \bar{X} + \sqrt{(b-a)^2 \ln(2/\alpha)/n}\right] > 1-\alpha \,.$$

非正态分布的区间估计

- 中心极限定理求枢轴量的近似分布, 设总体 X 的期望 $\mathbb{E}[X]=\mu$, 方差 $\mathbb{VAR}(X)=\sigma^2$, 有

$$W=\frac{\bar{X}-\mu}{\sigma\sqrt{n}}\xrightarrow{d}\mathcal{N}(0,1)\,.$$

  - 当 σ^2 已知时, 有

$$\Pr\left[-\mu_{\alpha/2}<\frac{\bar{X}-\mu}{\sigma\sqrt{n}}<\mu_{\alpha/2}\right]\approx 1-\alpha\,.$$

  - 当 σ^2 未知时, 用无偏样本方差 $S^2=\sum_{i=1}^n(X_i-\bar{X})^2/(n-1)$ 替代 σ^2, 于是有

$$\Pr\left[-\mu_{\alpha/2}<\frac{\bar{X}-\mu}{S/\sqrt{n}}<\mu_{\alpha/2}\right]\approx 1-\alpha\,.$$

区间估计：例 0.174

例 0.174 设 $X_1, X_2, \ldots, X_n$ 是来自总体 $X \sim B(p)$ 的样本, 求 p 的置信度为 $1-\alpha$ 的区间估计.

解答：例 0.174

解答:

- 利用集中不等式求解, 根据伯努利分布的性质有 $X_i \in \{0,1\}$ 以及 $p = \mathbb{E}[X]$, 根据切比雪夫不等式有

  $$\Pr[|\bar{X} - p| > \epsilon p] \leq 2\exp(-np\epsilon^2/3)\ ,$$

  设 $\alpha = 2\exp(-np\epsilon^2/3)$, 于是有

  $$\Pr[\bar{X} - \sqrt{3p\ln(2/\alpha)/n} < p < \bar{X} + \sqrt{3p\ln(2/\alpha)/n}] \geq 1 - \alpha\ ,$$

- 根据伯努利分布的性质有 $p = \mathbb{E}[X]$ 以及 $\mathbb{VAR}(X) = p(1-p)$, 利用中心极限定理可知

  $$W = \frac{n\bar{X} - np}{\sqrt{np(1-p)}} \xrightarrow{d} \mathcal{N}(0,1)\ .$$

  于是有

  $$\Pr\left[-\mu_{\alpha/2} < \frac{n\bar{X} - np}{\sqrt{np(1-p)}} < \mu_{\alpha/2}\right] \approx 1 - \alpha\ .$$

单侧置信区间

对某些实际问题, 我们往往只关心置信区间的上限或下限

- 例如, 次品率只关心上限, 产品的寿命只关心下限, 由此引入单侧置信区间及其估计.

定义 0.90 给定 $\alpha \in (0,1)$, 设样本 $X_1,\ldots,X_n$ 的统计量 $\hat{\theta}_1=\hat{\theta}_1(X_1,\ldots,X_n)$ 满足

$$\Pr[\theta > \hat{\theta}_1] \geq 1-\alpha\ ,$$

则称 $(\hat{\theta}_1,+\infty)$ 为 θ 的置信度为 $1-\alpha$ 的置信区间, $\hat{\theta}_1$ 称为单侧置信下限.

正态总体的单侧置信区间

对于正态总体, 可以将相关置信区间的估计都拓展到单侧置信估计.

- 若 $X_1, X_2, \ldots, X_n$ 是来自 $\mathcal{N}(\mu, \sigma^2)$ 的样本. 若 σ^2 已知, 讨论 μ 的置信度为 $1-\alpha$ 的单侧置信下限和上限.
- 构建枢轴量为

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

- 根据单侧置信上下限的定义有

$$\Pr\left[\frac{\bar{X} - \mu}{S/\sqrt{n}} < \mu_\alpha\right] = 1 - \alpha, \quad \Pr\left[\frac{\bar{X} - \mu}{S/\sqrt{n}} > -\mu_\alpha\right] = 1 - \alpha\ .$$

单侧置信区间估计：例 0.175

例 0.175 从一批出厂的灯泡中随机抽取 10 盏灯泡, 测试其寿命分别为: 1000, 1500, 1250, 1050, 950, 1000, 1150, 1050, 950, 1000 (单位：小时). 假设这批灯泡的寿命服从正态分布, 求这批灯泡平均寿命的置信度为 95% 的单侧置信下限.

解答：例 0.175

题目: 从一批出厂的灯泡中随机抽取 10 盏灯泡, 测试其寿命分别为: 1000, 1500, 1250, 1050, 950, 1000, 1150, 1050, 950, 1000 (单位: 小时). 假设这批灯泡的寿命服从正态分布, 求这批灯泡平均寿命的置信度为 95% 的单侧置信下限.

解答:

- 按照题意, 判断此例是求 σ 未知时 μ 的单侧置信区间, 枢轴变量选为 $W = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$. 计算样本均值及样本方差 (无偏) 分别为

$$\bar{X} = \sum_{i=1}^{10} X_i/10 = 1090 \quad \text{和} \quad S^2 = \sum_{i=1}^{10} \left(X_i - \bar{X}\right)^2/9 = 8800/3.$$

于是有

$$\Pr\left[\frac{\bar{X}-\mu}{S/\sqrt{10}} < t_{0.05}(9)\right] = 0.95,$$

查表可知 $t_{0.05}(9) = 1.833$ 可得

$$\mu > \bar{X} - t_{0.05}(9)S/\sqrt{10} = 1090 - \sqrt{8800/9} \times 1.833/\sqrt{10} > 1072.$$