第六章 参数估计

Instructor: 郝壮

haozhuang@buaa.edu.cn
School of Economics and Management
Beihang University

March 22, 2022

统计推断

回顾： 上一章: 常用统计量及其抽样分布, 目的在于对感兴趣的问题进行统计推断.

根据样本信息, 推断总体的统计规律, 这是统计推断的任务.

统计推断的基本问题：

- **估计问题：** 根据样本信息, 对总体分布中未知参数进行估计(本章重点)
- **假设检验** (第七章)

参数估计问题的提出

- 分布中所含的未知参数θ: 两点分布$b(1,p)$, 正态分布$N(\mu,\sigma^2)$.
- 分布中所含的未知参数的函数：服从正态分布$N(\mu,\sigma^2)$的变量不超过a的概率$P(X \leqslant a) = \Phi((a-\mu)/\sigma)$是未知参数$\mu,\sigma^2$的函数
- 分布的各种特征函数是未知参数：均值, 方差等

§6 参数估计

一般常用θ表示参数, 参数θ所有可能取值组成的集合称为**参数空间**, 常用Θ表示. 参数估计问题就是根据样本对上述各种未知参数作出估计.

参数估计的形式有两种：**点估计与区间估计**.

本章学习基本要求： 熟悉掌握两种参数估计形式的基本思想, 构造方法, 统计性质及其在不同分布情形下的计算, 软件(如STATA, R)实现

- §6.1 点估计与无偏性
- §6.2 矩估计与相合性
- §6.3 最大似然估计与EM算法(EM算法略)
- §6.4 最小方差无偏估计(略)
- §6.5 贝叶斯估计(略)
- §6.6 区间估计

§6.1 点估计(point estimate)的概念与无偏性(unbiasedness)

定义6.1.1 设 $x_1, x_2, ..., x_n$ 是来自总体X的一个样本, 我们用一个统计量 $\hat{\theta} = \hat{\theta}(x_1, x_2, ..., x_n)$ 的取值作为θ的估计值, 称为θ的**点估计**(point estimator), 简称**估计**(estimator).

说明:

- 因为样本 $x_1, x_2, ..., x_n$ 是随机变量(也可大写), θ的点估计(estimator) $\hat{\theta}$是一个随机变量. 如 $\hat{\mu} = \bar{x}$或$\hat{\mu} = \bar{X}$
- 若$x_1, x_2, ..., x_n$表示的是样本的观测值, 则$\hat{\theta} = \hat{\theta}(x_1, x_2, ..., x_n)$ 是θ的估计值(estimate).
- Estimator是对参数估计的random variable; estimate是该estimator的一个具体值.

在这里如何构造统计量$\hat{\theta}$并没有明确的规定, 只要它满足一定的合理性即可. 这就涉及到两个问题:

- 其一是如何给出估计, 即估计的方法问题;
- 其二是如何对不同的估计进行评价, 即估计的好坏判断标准.
  - 无偏性(Unbiasedness)
  - 有效性(Efficiency)
  - 相合性/一致性(Consistency)

无偏性(unbiasedness)

定义6.1.2 设 $\hat{\theta}=\hat{\theta}(x_1,x_2,...,x_n)$ 是θ的一个估计, θ 的参数空间为Θ, 若对任意的$\theta\in\Theta$有

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

则称$\hat{\theta}$是θ 的**无偏估计**(unbiased estimator), 否则称为**有偏估计**.

课堂练习： 设总体X分布未知, 总体的一阶和二阶矩存在, 记$E(X)=\mu, Var(X)=\sigma^2$, 证明：样本均值$\bar{x}$和样本方差$s^2$分别是$\mu$和$\sigma^2$的无偏估计.

证明:

课堂练习: 设总体X分布未知, 总体的一阶和二阶矩存在, 记$E(X)=\mu, Var(X)=\sigma^2$, 证明: 样本均值$\bar{x}$和样本方差$s^2$分别是$\mu$和$\sigma^2$的无偏估计.

证明: 因为$x_1,\ldots,x_n$与X同分布, 所以

$$\mathrm{E}(\bar{x})=\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}x_i\right)=\frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\left(x_i\right)=\frac{1}{n}\sum_{i=1}^{n}\mu=\mu$$

故$\bar{x}$是μ的无偏估计.

对任一总体而言, 样本均值是总体均值的无偏估计. 当总体k阶矩存在时, 样本k阶原点矩a_k是总体k阶原点矩μ_k的无偏估计. 证明同上.

$$
\begin{aligned}
E\left(s^2\right) &= E\left(\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2\right)\\
&= \frac{1}{n-1}E\left\{\sum_{i=1}^{n}[(x_i-\mu)-(\bar{x}-\mu)]^2\right\}\\
&= \frac{1}{n-1}E\left\{\sum_{i=1}^{n}(x_i-\mu)^2-2(\bar{x}-\mu)\sum_{i=1}^{n}(x_i-\mu)+n(\bar{x}-\mu)^2\right\}\\
&= \frac{1}{n-1}E\left\{\sum_{i=1}^{n}(x_i-\mu)^2-n(\bar{x}-\mu)^2\right\}\\
&= \frac{1}{n-1}\left\{\sum_{i=1}^{n}Var\left(x_i\right)-nVar(\bar{x})\right\}\\
&= \frac{1}{n-1}\left(n\sigma^2-n\times\frac{\sigma^2}{n}\right)=\sigma^2
\end{aligned}
$$

可以证明样本方差$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$不是总体方差$\sigma^2$的无偏估计, 因为 $E\left(s_n^2\right) = \frac{n-1}{n}\sigma^2$. 对此, 有如下两点说明:

(1) 当样本量趋于无穷时, 有$E\left(s_n^2\right) = \frac{n-1}{n}\sigma^2 \to \sigma^2$, 我们称$s_n^2$为$\sigma^2$的**渐近无偏估计**(asymptotically unbiased estimator).

(2) 若对s_n^2 作如下修正:

$$s^2 = \frac{ns_n^2}{n-1} = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$$

则 s^2是总体方差的无偏估计.

总结

(1)对任一总体而言, 样本均值是总体均值的无偏估计; 当总体k阶矩存在时, 样本k阶原点矩是总体k阶原点矩的无偏估计. k阶中心距则不一样.

(2)无偏性**不具有不变性**: 即若$\hat{\theta}$是θ的无偏估计, 一般而言, $g(\hat{\theta})$不是$g(\theta)$的无偏估计, 除非$g(\theta)$是θ的线性函数.

教材例6.1.2 虽然样本方差s^2是总体方差σ^2的无偏估计, 但样本标准差s不是σ的无偏估计, 而是渐近无偏估计.

例 6.1.2 设总体为 $N\left(\mu,\sigma^2\right), x_1,\cdots,x_n$ 是样本, 我们已经指出 s^2 是 σ^2 的无偏估计.下面来考察 s 是否是 σ 的无偏估计.由定理 **5.4.1**, $Y=\frac{(n-1)s^2}{\sigma^2}\sim\chi^2(n-1)$, 其密度函数为

$$p(y)=\frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}y^{\frac{n-1}{2}-1}\mathrm{e}^{-\frac{y}{2}},\quad y>0$$

从而

$$\begin{aligned}E\left(Y^{1/2}\right)&=\int_0^{\infty}y^{1/2}p(y)\mathrm{d}y\\&=\frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}\int_0^{\infty}y^{\frac{n}{2}-1}\mathrm{e}^{-\frac{y}{2}}\,\mathrm{d}y\\&=\frac{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)}=\sqrt{2}\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\end{aligned}$$

又因为$Y=\frac{(n-1)s^2}{\sigma^2}$, 所以有

$$E(s)=\frac{\sigma}{\sqrt{n-1}}E\left(Y^{1/2}\right)=\sqrt{\frac{2}{n-1}}\cdot\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}\cdot\sigma\equiv\frac{\sigma}{c_n}.$$

这说明 s 不是 σ 的无偏估计, 利用修正技术可得 $c_n\cdot s$ 是 σ 的无偏估计, 其中

$$c_n=\sqrt{\frac{n-1}{2}}\cdot\frac{\Gamma((n-1)/2)}{\Gamma(n/2)}$$

是修偏系数, 教材表 6.1.1 给出了 c_n 的部分取值.

可以证明,当 $n\to\infty$ 时有 $c_n\to 1$, 这说明 s 是 σ 的渐近无偏估计, 从而在样本容量较大时, 不经修正的 s 也是 σ 的一个很好的估计 ($n>30$时, s与σ的偏差小于1% (**教材表 6.1.1**)).

6.2.3 有效性 (efficiency)

定义6.2.3(有效性定义) 设$\hat{\theta}_1, \hat{\theta}_2$是$\theta$的两个无偏估计, 如果对任意的$\theta \in \Theta$, 有

$$\mathrm{Var}\left(\hat{\theta}_1\right) \leq \mathrm{Var}\left(\hat{\theta}_2\right)$$

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

- 如果存在方差最小(最有效)的无偏估计, 则有如下定义.
  定义6.4.2 设$\hat{\theta}$为θ的一个无偏估计, 如果对任意一个无偏估计$\tilde{\theta}$,对$\theta \in \Theta$, 上都有

  $$\mathrm{Var}(\widehat{\boldsymbol{\theta}}) \leq \mathrm{Var}(\widetilde{\boldsymbol{\theta}})$$

  则称$\hat{\theta}$为θ的**一致最小方差无偏估计**(Uniformly Minimum Variance Unbiased Estimator), 简记UMVUE.

例6.1.5 设 $x_1, x_2, ..., x_n$是取自某总体的样本, 记总体均值为μ, 总体方差为σ^2, 则$\hat{\mu}_1 = x_1, \hat{\mu}_2 = \bar{x}$ 都是μ的无偏估计, 但

$$\mathrm{Var}\left(\hat{\mu}_1\right) = \sigma^2, \quad \mathrm{Var}\left(\hat{\mu}_2\right) = \sigma^2/n$$

显然, 只要 $n > 1$, $\hat{\mu}_2$ 比$\hat{\mu}_1$有效. 这表明用全部数据的平均估计总体均值要比只使用部分数据更有效.

例6.1.6 设$x_1, \dots, x_n$是来自均匀总体$U(0,\theta)$中的样本. 人们常用最大观测值$x_{(n)}$来估计θ (这是θ的极大似然估计). 由于$E(x_{(n)}) = \dfrac{n}{n+1}\theta$ (教材例6.2.5), 所以$x_{(n)}$不是θ 的无偏估计, 而是θ 的渐近无偏估计. 经过修偏后可以得到θ 的一个无偏估计: $\hat{\theta}_1 = \frac{n+1}{n}x_{(n)}$. 且

$$\mathrm{Var}\left(\hat{\theta}_1\right) = \left(\frac{n+1}{n}\right)^2 \mathrm{Var}\left(x_{(n)}\right) = \left(\frac{n+1}{n}\right)^2 \frac{n}{(n+1)^2(n+2)}\theta^2 = \frac{\theta^2}{n(n+2)}$$

另一方面, 由于总体均值为$\theta/2$, 可以使用样本均值估计总体均值, 于是可得到θ的另一个无偏估计$\hat{\theta}_2 = 2\bar{x}$(这是$\theta$的矩估计), 且

$$\mathrm{Var}\left(\hat{\theta}_2\right) = 4\,\mathrm{Var}(\bar{x}) = \frac{4}{n}\mathrm{Var}(X) = \frac{4}{n}\cdot\frac{\theta^2}{12} = \frac{\theta^2}{3n}$$

由此, 当$n > 1$时, $\hat{\theta}_1$比 $\hat{\theta}_2$ 有效.

§6.2 矩估计(method of moment)及相合性/一致性(consistency)

矩估计和替换方法的提出:

- 皮尔逊 1900 (Carl Pearson 1857-1936)

6.2.1 替换原理和矩法估计

替换原理是指用**样本矩及其函数**去替换相应的**总体矩及其函数**, 譬如:

- 用样本均值$\bar{x}$(样本的一阶原点矩)估计总体均值$E(X)$

矩法估计的实质是用经验分布函数去替换总体分布, 其理论基础是格利文科定理.

另外把事件出现的概率看作是$0, 1$变量中1出现的比例, 则频率也是样本的一阶原点矩, 所以可以

- 用事件A出现的频率估计事件A发生的概率
- 用样本的p分位数估计总体的p分位数(将p分位数看作0-1变量出现和不出现)
- 用样本中位数估计总体中位数(50分位数)

由于样本的中心距(样本方差是二阶中心矩)都是原点矩的函数. 所以可以

- 用样本方差s_n^2估计总体方差$Var(X)$

注意： 教材中的表达是

- 用样本方差s^2估计总体方差$Var(X)$

这是近似表达(**严格上不符合矩估计的定义**). 但后文中会学到, s^2和s_n^2都是$Var(X)$的一致估计, 故教材中用s^2作为$Var(X)$的矩估计.

例6.2.1 对某型号的20辆汽车记录其每5L汽油的行驶里程(km), 观测数据如下:

29.8	27.6	28.3	27.9	30.1	28.7	29.9	28.0
27.9	28.7	28.4	27.2	29.5	28.5	28.0	30.0
29.1	29.8	29.6	26.9				

经计算有

$$\bar{x} = 28.695, \quad s^2 = 0.9668, \quad m_{0.5} = 28.6$$

由此给出总体均值, 方差和中位数的估计分别为: 28.695, 0.9668 和 28.6.

概率函数$P(x,\theta)$已知时, 未知参数的矩估计

设总体具有已知的概率函数 $P(x;\theta_1,\dots,\theta_k)$, $x_1,x_2,\dots,x_n$ 是样本, 假定总体的k阶原点矩μ_k存在, 若$\theta_1,\dots,\theta_k$能够表示成$\mu_1,\dots,\mu_k$ 的函数$\theta_j=\theta_j(\mu_1,\dots,\mu_k)$, 则可给出诸$\theta_j$的矩法估计为

$$\hat{\theta}_j=\theta_j\left(a_1,\cdots,a_k\right),\quad j=1,\cdots,k$$

其中

$$a_j=\frac{1}{n}\sum_{i=1}^{n}x_i^j$$

是前k阶样本原点矩.

(样本矩的函数可以估计总体矩的函数. **拓展知识:** 底层原理: Slusky Theorem)

例6.2.2 设总体服从指数分布, 其密度函数为

$$p(x;\lambda) = \lambda \mathrm{e}^{-\lambda x}, \quad x > 0$$

$x_1, \ldots, x_n$是样本, 由于只有一个未知参数λ, 故$k = 1$. 对于总体, 由于$E(X) = 1/\lambda$, 即$\lambda = 1/E(X)$, 故λ的矩法估计为

$$\hat{\lambda} = 1/\bar{x}$$

另外, 由于$Var(X) = 1/\lambda^2$, 其反函数为$\lambda = 1/\sqrt{\mathrm{Var}(X)}$. 因此, 从替换原理来看, λ的矩法估计也可取为

$$\hat{\lambda}_1 = 1/s$$

s **为样本标准差**.

这说明**矩估计可能是不唯一的**! 这是矩法估计的一个缺点, 此时通常应该尽量**采用低阶矩**给出未知参数的估计.

思考:

- 为什么可以用s替换$\sqrt{\mathrm{Var}(X)}$? (未来会学到Slusky Theorem, plim operates through continuous functions, 现在仅需知道样本矩的函数可以估计总体矩的函数)
- 上面用s替换$\sqrt{\mathrm{Var}(X)}$符合矩估计的定义么?

(中心距是原点矩的函数. 严格意义上应使用s_n)

例6.2.3 $x_1, x_2, \ldots, x_n$是来自均匀分布$U(a,b)$的样本, a与b均是未知参数, 这里$k=2$, 由于

$$E(X) = \frac{a+b}{2}, \quad \mathrm{Var}(X) = \frac{(b-a)^2}{12}$$

不难推出

$$a = E(X) - \sqrt{3\,\mathrm{Var}(X)}, \quad b = E(X) + \sqrt{3\,\mathrm{Var}(X)}$$

由此即可得到a,b的矩估计:

$$\hat{a} = \bar{x} - \sqrt{3}s, \quad \hat{b} = \bar{x} + \sqrt{3}s$$

(严格意义上应使用s_n)

概率函数$P(x, \theta)$未知时，未知参数的矩估计

例： 设总体X的均值μ和方差σ^2都存在且未知. $x_1, \ldots, x_n$是取自总体X的一个样本. 求均值μ和方差σ^2的矩估计.

解：

概率函数$P(x,\theta)$未知时，未知参数的矩估计

例： 设总体X的均值μ和方差σ^2都存在且未知. $x_1,\ldots,x_n$是取自总体X的一个样本. 求均值μ和方差σ^2的矩估计.

解： 先求总体矩:

$$\begin{aligned} \mu_1 &= E(X) = \mu \\ \mu_2 &= E\left(X^2\right) = \operatorname{Var}(X) + E^2(X) = \sigma^2 + \mu^2 \end{aligned}$$

再求样本矩:

$$\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}, \quad \hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

所以:

$$\begin{cases} \hat{\mu} = \bar{x} \\ \hat{\sigma}^2 = \hat{\mu}_2 - \hat{\mu}_1^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \end{cases}$$

正是s_n^2的表达式，它比 s^2 略小

§6.2.3 点估计的评价标准之一:相合性/一致性(consistency)

如果有足够的观测值, 根据格利文科定理, 随着样本量的不断增大, 经验分布函数逼近真实分布函数, 因此可以要求估计量随着样本量的不断增大而逼近参数真值, 这就是相合性.

定义6.2.1 设$\theta \in \Theta$为未知参数, $\hat{\theta}_n = \hat{\theta}_n(x_1, \cdots, x_n)$是$\theta$的一个估计量, n 是样本容量, 若对任何一个$\epsilon > 0$, 有 $\hat{\theta}_n$依概率收敛到θ

$$\lim_{n\to\infty} P\left(\left|\hat{\theta}_n - \theta\right| > \varepsilon\right) = 0$$

或$plim(\hat{\theta}_n) = \theta$ ($\hat{\theta}_n$的概率极限是θ).

则称$\hat{\theta}_n$为θ参数的**相合(或一致)估计**(consistent estimator).

注意

- 相合性被认为是对估计的一个最基本要求, 如果一个估计量, 在样本量不断增大时, 它都不能把被估参数估计到任意指定的精度, 那么这个估计是很值得怀疑的.
- 通常, 不满足相合性要求的估计量一般不予考虑. (拓展：但不满足无偏性要求的估计量在大样本时经常被使用, 如：工具变量估计量)
- 证明估计的相合性一般可应用大数定律或直接由定义来证.
- 若把依赖于样本量n的估计量$\hat{\theta}_n$看作一个随机变量序列, 相合性就是$\hat{\theta}_n$依概率收敛于θ, 所以证明估计的相合性可应用依概率收敛的性质及各种大数定律.

复习：辛钦大数定律 (Khinchin's LLN)

定理4.2.4 若随机变量序列$\{X_n\}$独立同分布, 且X_i的数学期望存在, 则$\{X_n\}$服从大数定律, 即对任意 $\varepsilon > 0$, 有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right)\right| < \varepsilon\right) = 1$$

- 辛钦大数定律要求独立同分布, 不要求方差存在(可为无穷大).

复习：辛钦大数定律 (Khinchin's LLN)

辛钦大数定律推论1:

- 对随机变量X独立观察n次, 记每次观察值为X_i, 当n足够大时, 可以用$\frac{1}{n}\sum_{i=1}^{n} X_i$近似$E(X)$.

辛钦大数定律推论2:

- 若$E(X^k)$存在, 对随机变量X独立观察n次, 记每次观察值为X_i, 当n足够大时, 可以用$\frac{1}{n}\sum_{i=1}^{n} X_i^k$近似$E(X^k)$.
(因为若随机变量序列$\{X_n\}$独立同分布, 则$\{X_n^k\}$独立同分布)

在判断估计的相合性时有两个定理是很有用的.

定理6.2.1 设 $\hat{\theta}_n = \hat{\theta}_n (x_1, \cdots, x_n)$是$\theta$的一个估计量, 若

$$\lim_{n\to\infty} E\left(\hat{\theta}_n\right) = \theta, \quad \lim_{n\to\infty} \mathrm{Var}\left(\hat{\theta}_n\right) = 0$$

则$\hat{\theta}_n$是θ的相合估计.

定理6.2.1 设 $\hat{\theta}_n = \hat{\theta}_n(x_1, \cdots, x_n)$是$\theta$的一个估计量, 若

$$\lim_{n\to\infty} E\left(\hat{\theta}_n\right) = \theta, \quad \lim_{n\to\infty} \mathrm{Var}\left(\hat{\theta}_n\right) = 0$$

则$\hat{\theta}_n$是θ的相合估计.

证明: 对任意的 $\varepsilon > 0$, 由切比雪夫不等式有

$$P\left(\left|\hat{\theta}_n - E\hat{\theta}_n\right| \geqslant \frac{\varepsilon}{2}\right) \leqslant \frac{4}{\varepsilon^2} \mathrm{Var}\left(\hat{\theta}_n\right).$$

另一方面, 由 $\lim_{n\to\infty} E\left(\hat{\theta}_n\right) = \theta$ 可知, 当 n 充分大时有

$$\left|E\hat{\theta}_n - \theta\right| < \frac{\varepsilon}{2}.$$

注意到此时如果 $\left|\hat{\theta}_n - E\hat{\theta}_n\right| < \frac{\varepsilon}{2}$ (定理条件加切比雪夫不等式可以保证), 就有

$$\left|\hat{\theta}_n - \theta\right| \leqslant \left|\hat{\theta}_n - E\hat{\theta}_n\right| + \left|E\hat{\theta}_n - \theta\right| < \varepsilon,$$

故

$$\left\{\left|\hat{\theta}_n - E\hat{\theta}_n\right| < \frac{\varepsilon}{2}\right\} \subset \left\{\left|\hat{\theta}_n - \theta\right| < \varepsilon\right\}$$

等价地

$$\left\{\left|\hat{\theta}_n - E\hat{\theta}_n\right| \geqslant \frac{\varepsilon}{2}\right\} \supset \left\{\left|\hat{\theta}_n - \theta\right| \geqslant \varepsilon\right\}$$

由此即有, 当$n \to \infty$时

$$P\left(\left|\hat{\theta}_n - \theta\right| \geqslant \varepsilon\right) \leqslant P\left(\left|\hat{\theta}_n - E\hat{\theta}_n\right| \geqslant \frac{\varepsilon}{2}\right) \leqslant \frac{4}{\varepsilon^2} \mathrm{Var}\left(\hat{\theta}_n\right) \to 0,$$

定理得证.

例6.2.5 设 $x_1, x_2, \ldots, x_n$ 是来自均匀总体$U(0,\theta)$的样本，证明$x_{(n)}$是θ相合估计.

证明： 由次序统计量的分布，我们知道 $\hat{\theta} = x_{(n)}$ 的概率密度函数为

$$p(y) = ny^{n-1}/\theta^n, \quad y < \theta$$

故有

$$E(\hat{\theta}) = \int_0^\theta ny^n \mathrm{d}y/\theta^n = \frac{n}{n+1}\theta \to \theta(n \to \infty)$$
$$E(\hat{\theta}^2) = \int_0^\theta ny^{n+1} \mathrm{d}y/\theta^n = \frac{n}{n+2}\theta^2$$
$$\mathrm{Var}(\hat{\theta}) = \tfrac{n}{n+2}\theta^2 - \left(\tfrac{n}{n+1}\theta\right)^2 = \tfrac{n}{(n+1)^2(n+2)}\theta^2 \to 0 \quad (n \to \infty)$$

由定理6.2.1可知，$x_{(n)}$是θ的相合估计.

连续映射定理

定理6.2.2 若 $\hat{\theta}_{n1}, \cdots, \hat{\theta}_{nk}$分别是$\theta_1, \ldots, \theta_k$的相合估计，$\eta = g(\theta_1, \ldots, \theta_k)$是$\theta_1, \ldots, \theta_k$的连续函数，
则$\hat{\eta}_n = g\left(\hat{\theta}_{n1}, \cdots, \hat{\theta}_{nk}\right)$ 是η的相合估计.

拓展知识: 定理6.2.2又叫连续映射定理(continuous mapping theorem, CMT), Slutsky theorem 是CMT的特殊形式.

证明见教材.

由**大数定律及定理6.2.2**可以推出： **矩估计一般都具有相合性**. 比如：

- 样本均值$\bar{x}$是总体均值μ的相合估计
- 样本方差s_n^2是总体方差σ^2的相合估计
- 样本方差s^2是总体方差σ^2的相合估计(参数的相合估计不止一个)
- 样本标准差s是总体标准差σ的相合估计

以下给出证明:

- 样本均值$\bar{x}$是总体均值μ的相合估计

证明： 设 $x_1, x_2, \ldots, x_n$ 是来自总体X的样本, $E(X) = \mu$. 由于$x_1, x_2, \ldots, x_n$ 独立同X分布, 则$E(x_i) = \mu, \forall i$, 故

$$\lim_{n \to +\infty} \frac{1}{n} \sum_{i=1}^{n} x_i = \mu$$

所以

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

- 样本方差s^2是总体方差σ^2的相合估计
- 样本方差s_n^2是总体方差σ^2的相合估计

证明： 设 $x_1, x_2, \ldots, x_n$ 是来自总体X的样本, $E(X) = \mu$.

由辛钦大数定律及其推论

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} x_i = \lim_{n\to+\infty} \bar{x} \to \mu$$

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} x_i^2 \to E\left(X^2\right)$$

用与辛钦大数定律相同证明方法, 下式亦成立

$$\lim_{n\to+\infty} \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 \to E\left(X^2\right)$$

故

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

所以：样本方差s_n^2是总体方差σ^2的相合估计

同理：样本方差s^2也是总体方差σ^2的相合估计

- 样本标准差s是总体标准差σ的相合估计

证明： 设 $x_1, x_2, \ldots, x_n$ 是来自总体X的样本, $E(X) = \mu$.

由于：样本方差s^2是总体标准差σ^2的相合估计

和定理6.2.2 若 $\hat{\theta}_{n1}, \cdots, \hat{\theta}_{nk}$分别是$\theta_1, \ldots, \theta_k$的相合估计,
$\eta = g(\theta_1, \ldots, \theta_k)$是$\theta_1, \ldots, \theta_k$的连续函数,
则$\hat{\eta}_n = g\left(\hat{\theta}_{n1}, \cdots, \hat{\theta}_{nk}\right)$ 是η的相合估计.

所以：样本方差s是总体标准差σ的相合估计

同理, 由**定理6.2.2** , 任何k 阶样本原点矩$\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$都是$k$阶总体原点矩$\mu_k = EX^k$的相合估计(且无偏).

小结

- 样本均值$\bar{x}$：是总体均值的相合, 无偏估计
- 样本方差s^2：是总体方差的相合, 无偏估计
- 样本标准差s：是总体标准差的相合, 有偏估计(渐近无偏)

例6.2.6 设一个试验有三种可能的结果, 发生的概率分别为

$$p_1 = \theta^2, \quad p_2 = 2\theta(1-\theta), \quad p_3 = (1-\theta)^2$$

现做n次试验, 观测到三种结果发生的次数分别为n_1, n_2, n_3, 可以采取频率替换方法估计θ. 由于可以有三个不同的θ的表达式

$$\theta = \sqrt{p_1}, \quad \theta = 1 - \sqrt{p_3}, \quad \theta = p_1 + p_2/2$$

从而可以给出θ三种不同的频率替换估计, 分别为

$$\hat{\theta}_1 = \sqrt{n_1/n}, \quad \hat{\theta}_2 = 1 - \sqrt{n_3/n}, \quad \hat{\theta}_3 = (n_1 + n_2/2)/n$$

由大数定律, $n_1/n, n_2/n, n_3/n$ 分别是p_1, p_2, p_3的相合估计. 由定理6.2.2, 上述三个估计都是θ的相合估计.

思考： 用频率代替概率如何体现了"矩估计"的思想?

作业2

6.1课后习题： 1, 3, 6, 7

6.2课后习题： 1, 3, 4, 5

6.3 极(最)大似然估计 (maximum likelihood estimation)

- 高斯(1821)和费希尔(1922)
- 极大似然估计通过观察样本的结果出现的可能(似然)推断总体
- 需要总体的分布已知 (强假设, 某些问题难以满足)
- 一般需要样本量较大, 因此不讨论无偏性
- 下面通过两个例子叙述最大似然估计的基本思想

例6.3.1a 设有外形完全相同的两个箱子, 甲箱中有99个白球和1个黑球, 乙箱中有99个黑球和1个白球, 今随机地抽取一箱, 并从中随机抽取一球, 结果取得白球. 问这球是从哪一个箱子中取出?

解:

例6.3.1a 设有外形完全相同的两个箱子, 甲箱中有99个白球和1个黑球, 乙箱中有99个黑球和1个白球, 今随机地抽取一箱, 并从中随机抽取一球, 结果取得白球. 问这球是从哪一个箱子中取出?

解: 由于甲箱中抽出白球的概率为0.99, 乙箱中抽出白球的概率为0.01, 因此最像(最大似然)从甲箱抽出.

例6.3.1a 设有外形完全相同的两个箱子, 甲箱中有99个白球和1个黑球, 乙箱中有99个黑球和1个白球, 今随机地抽取一箱, 并从中随机抽取一球, 结果取得白球. 问这球是从哪一个箱子中取出?

解: 由于甲箱中抽出白球的概率为0.99, 乙箱中抽出白球的概率为0.01, 因此最像(最大似然)从甲箱抽出.

例6.3.1b 设有外形完全相同的两个箱子, 甲箱中白球比例为p_1, 乙箱中白球比例为p_2, 已知$p_1 > p_2$. 今随机地抽取一箱, 并从中随机抽取一球, 结果取得白球. 问该球最可能从哪个箱中取出?

例6.3.2 设产品分为合格和不合格品两类. 用随机变量X表示某个产品经检查的不合格数, 则$X=0$表示合格品, $X=1$表示不合格品, 则X服从两点分布$b(1,p)$, 其中p未知.

现抽取n个产品看是否合格, 得到样本$x_1,...,x_n$, 这批观测值发生的概率为

$$\begin{aligned} P\left(X_1=x_1,\cdots,X_n=x_n;p\right) &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \\ &= p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i} \end{aligned}$$

由于p是未知的, 根据最大似然原理, 我们应选择p使得上式表示的概率尽可能大. 由于$x_1,...,x_n$可观察, 可将上式看作是未知参数p的函数, 用$L(p)$表示, 称作**似然函数**(likelihood function),

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

似然函数(likelihood function),

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

接下来的任务是找出p, 使得上式最大.

一般做法是将似然函数取对数, 并令其一阶导等于0.

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

解得p的最大似然估计为

$$\hat{p}=\hat{p}\left(x_1,\cdots,x_n\right)=\sum_{i=1}^{n}x_i/n=\bar{x}$$

思考： p的矩估计是什么?

6.3 极(最)大似然估计

对离散总体, 设有样本观测值$x_1, ..., x_n$. 我们可写出该观测值出现的**概率** (依赖于未知参数θ). 将该**概率**看成是θ的函数, 用$L(\theta)$表示, 即为似然函数.

$$L(\theta) = P\left(X_1 = x_1, \cdots, X_n = x_n; \theta\right)$$

极大似然估计就是找θ的估计值$\hat{\theta}$使得上式最大.

对连续总体, 样本观测值$x_1, ..., x_n$出现的概率均为0, 因此需要用联合概率密度函数表示其似然函数. 有如下定义:

定义6.3.1 设总体的概率函数为$p(x;\theta)$, $\theta\in\Theta$是参数θ可能取值的参数空间, $x_1,x_2,\dots,x_n$ 是样本, 将样本的联合概率函数看成θ的函数, 用$L(\theta;x_1,x_2,\dots,x_n)$表示, 简记为$L(\theta)$,

$$L(\theta)=L\left(\theta;x_1,\cdots,x_n\right)=p\left(x_1;\theta\right)\cdot p\left(x_2;\theta\right)\cdots p\left(x_n;\theta\right)$$

$L(\theta)$称为样本的**似然函数**.

如果某统计量$\hat{\theta}=\hat{\theta}\left(x_1,\cdots,x_n\right)$满足

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

则称$\hat{\theta}$是θ的极(最)大似然估计, 简记为MLE(Maximum Likelihood Estimate).

- 人们通常更习惯于由对数似然函数$lnL(\theta)$出发寻找θ的极大似然估计. 当$L(\theta)$是可微函数时, 求导是求极大似然估计最常用的方法, 对$lnL(\theta)$求导更加简单些.

例6.3.3 设一个试验有三种可能的结果，发生的概率分别为

$$p_1=\theta^2,\quad p_2=2\theta(1-\theta),\quad p_3=(1-\theta)^2$$

现做n次试验，观测到三种结果发生的次数分别为n_1,n_2,n_3 $(n_1+n_2+n_3=n)$.

例6.2.6 给出了θ的三个矩估计，在此给出最大似然估计

解： 似然函数为

$$\begin{aligned} L(\theta) &= \left(\theta^2\right)^{n_1}[2\theta(1-\theta)]^{n_2}\left[(1-\theta)^2\right]^{n_3} \\ &= 2^{n_2}\theta^{2n_1+n_2}(1-\theta)^{2n_3+n_2} \end{aligned}$$

其对数似然函数为

$$\ln L(\theta)=(2n_1+n_2)\ln\theta+(2n_3+n_2)\ln(1-\theta)+n_2\ln 2$$

将之关于θ求导, 并令其为0得到似然方程

$$\frac{2n_1+n_2}{\theta}-\frac{2n_3+n_2}{1-\theta}=0$$

解得

$$\hat{\theta}=\frac{2n_1+n_2}{2\left(n_1+n_2+n_3\right)}=\frac{2n_1+n_2}{2n}$$

由于

$$\frac{\partial^2\ln L(\theta)}{\partial\theta^2}=-\frac{2n_1+n_2}{\theta^2}-\frac{2n_3+n_2}{(1-\theta)^2}<0$$

所以$\hat{\theta}$是极大值点.

例6.3.4 对正态总体$N(\mu,\sigma^2)$, $\theta=(\mu,\sigma^2)$是二维参数, 设有样本 $x_1,x_2,\ldots,x_n$, 则似然函数及其对数分别为

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

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

课堂练习: 利用对数似然函数$\ln L\left(\mu,\sigma^2\right)$推导出$\mu$和$\sigma^2$的MLE.

解： 将$\ln L\left(\mu,\sigma^2\right)$分别关于两个分量求偏导并令其为0, 即得到似然方程组

$$\frac{\partial \ln L\left(\mu,\sigma^2\right)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0$$

$$\frac{\partial \ln L\left(\mu,\sigma^2\right)}{\partial \sigma^2} = \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 - \frac{n}{2\sigma^2} = 0$$

解此方程组, 可得μ, σ^2 的极大似然估计为

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

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

利用二阶导函数矩阵的非正定性可以说明上述估计使得似然函数取极大值.

虽然求导函数是求极大似然估计最常用的方法，但并不是在所有场合求导都是可行的.

例6.3.5 设$x_1, x_2, \ldots, x_n$是来自均匀总体$U(0, \theta)$的样本，试求θ的极大似然估计.

解: 似然函数

$$L(\theta) = \frac{1}{\theta^n} \prod_{i=1}^{n} I_{\{0<x_i\leq\theta\}}$$

(教材错误(不影响推导)：教材进一步写出 $= \frac{1}{\theta^n} I_{\left\{x_{(n)}\leq\theta\right\}}$)

$x_{(n)}$是最大次序统计量.

要使$L(\theta)$达到最大，首先示性函数(指示函数，indicator function)取值应为1，其次是$1/\theta^n$尽可能大. 由于$1/\theta^n$是θ的单调减函数，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：$\hat{\theta} = x_{(n)}$.

极大似然估计的不变性 (invariance of MLE)

极大似然估计有一个简单而有用的性质：如果$\hat{\theta}$是θ的极大似然估计，则对任一**函数** $g(\theta)$，其极大似然估计为$g(\hat{\theta})$. 该性质称为极大似然估计**的不变性**，从而使一些复杂结构的参数的极大似然估计的获得变得容易了.

证明为拓展内容，见下页.

极大似然估计的不变性 (invariance of MLE)

极大似然估计的不变性: 设 $\hat{\theta}$是θ的MLE (原极大似然估计问题). 则对于任何函数 g, $\hat{\eta} = g(\hat{\theta})$ 是$\eta = g(\theta)$ 的 MLE.

证明: g是1-1映射(one-to-one): 首先, 要求$\eta = g(\theta)$的最大似然估计, 应该构造一个以η为变量的似然函数.

因为$\eta = g(\theta)$, 所以$\theta = g^{-1}(\eta)$. 在原似然函数可写为$L(\theta) = L\left(g^{-1}(\eta)\right)$. $L\left(g^{-1}(\eta)\right)$即即为以$\eta$为变量的似然函数变表达$L(\eta; x_1, ..., x_n)$.

由原问题已知, 这个似然函数在$\theta = g^{-1}(\eta) = \hat{\theta}$时取到极大值$L(\hat{\theta})$. 所以, 要选择$\hat{\eta}$以最大化$L(\theta) = L\left(g^{-1}(\eta)\right)$, 应选择$\hat{\eta}$使得$g^{-1}(\hat{\eta}) = \hat{\theta}$, 即, $\hat{\eta} = g(\hat{\theta})$.

如果g是m-1映射, 则似然函数在$\theta = g^{-1}(\eta) = \hat{\theta}$时依然取到极大值$L(\hat{\theta})$. QED.

例6.3.6 设 $x_1, x_2, \ldots, x_n$是来自正态总体$N(\mu, \sigma^2)$的样本, μ, σ^2的极大似然估计为

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

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

于是由不变性可得如下参数的极大似然估计:

- 总体标准差σ的MLE是$\hat{\sigma} = s_n$
- 概率$P(X < 3) = \Phi\left(\frac{3-\mu}{\sigma}\right)$的MLE是$\Phi\left(\frac{3-\bar{x}}{s_n}\right)$
- 总体0.90分位$x_{0.90} = \mu + \sigma u_{0.90}$的MLE是$\bar{x} + s_n \cdot u_{0.90}$，其中$u_{0.90}$为标准正态分布的0.90分位数.

课堂练习

甲, 乙两人下棋, 用 p 表示甲在每局中获胜的概率. 如果在5局中, 甲胜了3局, 问: p 的MLE? (做之前请先猜答案)

解:

课堂练习

甲, 乙两人下棋, 用 p 表示甲在每局中获胜的概率. 如果在5局中, 甲胜了3局, 问: p 的MLE? (做之前请先猜答案)

解: 似然函数:

$$L(p) = C_5^3 p^3 (1-p)^2$$

对$L(p)$ 求导数, 令一阶导等于0

$$L'(p) = C_5^3 \left[3p^2(1-p)^2 - 2p^3(1-p)\right]$$
$$= C_5^3 p^2 (1-p)[3(1-p) - 2p] = 0$$
$$p = 0, 1, \frac{3}{5}$$

$p = 0, p = 1$不能使甲获胜3局 (本质是只有$p = \frac{3}{5}$, 二阶导小于0), 故$\hat{p} = \frac{3}{5}$

课堂练习

总体服从指数分布$X \sim Exp(\lambda)$, 设有样本$x_1, ..., x_n$：求λ的MLE?

解：

课堂练习

总体服从指数分布$X \sim Exp(\lambda)$, 设有样本$x_1, ..., x_n$:
求λ的MLE?

解: 指数分布的概率密度函数为$f(x;\lambda) = \lambda e^{-\lambda x}, \quad x \geq 0.$
所以似然函数:

$$L(\lambda) = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} x_i\right)$$

对数似然函数:

$$\ln L(\lambda) = n \ln \lambda - \lambda \sum_{i=1}^{n} x_i$$

一阶条件: $\frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$, 故$\hat{\lambda} = \frac{1}{\bar{x}}$

拓展知识

未来接触到的复杂似然函数很多时候很难获得解析解, 一般采用数量解的方式给出MLE估计.

如何快速获得可信的全局最大值是数量解方法关注的问题, 本门课不会讲到.

最大似然估计的EM算法(略)

最大似然估计的EM算法是求解复杂似然函数问题的一种方法, 略.

6.3.3 最大似然估计的渐近正态性

最大似然估计的渐近正态性的证明较复杂, 是研究生学习内容.

结合**6.4.4**, 在此仅需掌握:

- 在较宽松的条件下(3个正则条件, regularity conditions), 若**总体分布形式已知**, 最大似然估计具有**相合性, 渐近正态性, 且是渐近最有效的**(达到克莱默-劳下界/$C-R$下界).

6.3.3 最大似然估计的渐近正态性

定理6.3.1 设总体X有密度函数$p(x;\theta)$, 若

- (1) 对任意的x, $\ln p(x;\theta)$对θ的前三阶导都存在(保证泰勒展开逼近和有限方差)
- (2) $p(x;\theta)$对θ的前三阶导都小于某些具有有限期望的关于x的函数(保证泰勒展开可以截取)
- (3) $\ln p(x;\theta)$对θ的前两阶导的期望都存在 (保证**费希尔信息量/信息矩阵**存在)

则θ的极大似然估计$\hat{\theta}$具有**相合性**和**渐近正态性**

$$\hat{\theta}_n \sim AN\left(\theta, \frac{1}{nI(\theta)}\right)$$

6.3.3 最大似然估计的渐近正态性

θ的极大似然估计θ具有相合性和渐近正态性

$$\hat{\theta}_n \sim AN\left(\theta, \frac{1}{nI(\theta)}\right)$$

其中

$$I(\theta) \equiv E\left[\frac{\partial}{\partial\theta}\ln p(x;\theta)\right]^2 = \int_{-\infty}^{\infty}\left(\frac{\partial \ln p}{\partial\theta}\right)^2 p(x;\theta)\mathrm{d}x$$

称为**费希尔信息量**. $I(\theta)$越大, 说明总体分布中包含未知参数θ的信息越多(进而MLE的方差越小).

$\frac{1}{nI(\theta)}$ 称为θ的无偏估计的方差的**克拉默-拉奥下界**(Cramo-Rao Lower Bound), **是任何估计方法得到的无偏估计**所能达到的**最小方差**.(故MLE是渐近最有效的).

一般形式： 定义6.4.3 (Cramer-Rao)不等式：
设$T = T(x_1, ..., x_n)$是$g(\theta)$的任何一个无偏估计, 称

$$\frac{[g'(\theta)]^2}{nI(\theta)}$$

为$g(\theta)$的无偏估计的方差的$C-R$下界, 简称$g(\theta)$的$C-R$下界. 进一步, 有以下不等式成立：

$$\mathrm{Var}(T) \geqslant [g'(\theta)]^2 / (nI(\theta))$$

证明略.

- 可以看出, $g(\theta) = \theta$的特殊情况下, $\frac{1}{nI(\theta)}$是$C-R$下界
- 可以看出, $C-R$下界可计算的前提条件是总体分布函数形式已知.

拓展练习1：求解MLE的渐近分布

例6.3.9 设$x_1, \ldots, x_n$是来自$N(\mu, \sigma^2)$的样本，可以验证该总体分布在σ^2已知或μ已知时，满足上述三个正则条件.

(1) 在σ^2已知时，μ的MLE为$\hat{\mu} = \bar{x}$，由**定理6.3.1**，$\hat{\mu}$服从渐近正态分布，下面求$I(\mu)$.

$$\ln p(x) = -\ln\sqrt{2\pi} - \frac{1}{2}\ln\sigma^2 - \frac{(x-\mu)^2}{2\sigma^2}$$

$$\frac{\partial \ln p}{\partial \mu} = \frac{x-\mu}{\sigma^2}$$

$$I(\mu) = E\left(\frac{x-\mu}{\sigma^2}\right)^2 = \frac{E(x-\mu)^2}{\sigma^4} = \frac{1}{\sigma^2}$$

从而有

$$\hat{\mu} \sim AN\left(\mu, \sigma^2/n\right)$$

这与$\hat{\mu}$的精确分布相同.

拓展练习1：求解MLE的渐近分布

(2) 在μ已知, σ^2的MLE为$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$, 下面求$I(\sigma^2)$.

$$\frac{\partial \ln p}{\partial \sigma^2}=-\frac{1}{2\sigma^2}+\frac{1}{2\sigma^4}(x-\mu)^2=\frac{(x-\mu)^2-\sigma^2}{2\sigma^4}$$

$$\begin{aligned}
I\left(\sigma^2\right) &= \frac{E\left[(x-\mu)^2-\sigma^2\right]^2}{4\sigma^8} \\
&= \frac{\operatorname{Var}\left((x-\mu)^2\right)}{4\sigma^8} \\
&= \frac{\operatorname{Var}\left(\frac{(x-\mu)^2}{\sigma^2}\times\sigma^2\right)}{4\sigma^8} \quad \text{(构造 } \frac{(x-\mu)^2}{\sigma^2}\sim\chi^2(1)\text{ , 方差为2)} \\
&= \frac{1}{2\sigma^4}
\end{aligned}$$

从而

$$\hat{\sigma}^2\sim AN\left(\sigma^2, 2\sigma^4/n\right)$$

拓展练习2：求解费希尔信息量矩阵

例6.4.4 设总体分布为泊松分布$P(\lambda)$, 其分布列为

$$p(x;\lambda)=\frac{\lambda^x}{x!}\mathrm{e}^{-\lambda},\quad x=0,1,\cdots$$

所以

$$\ln p(x;\lambda)=x\ln\lambda-\lambda-\ln(x!)$$
$$\frac{\partial}{\partial\lambda}\ln p(x;\lambda)=\frac{x}{\lambda}-1$$

所以

$$I(\lambda)=E\left(\frac{x-\lambda}{\lambda}\right)^2=\frac{1}{\lambda}$$

拓展练习3：求解费希尔信息量矩阵

例6.4.5 设总体分布为指数分布，其密度函数为

$$p(x;\theta)=\frac{1}{\theta}\exp\left\{-\frac{x}{\theta}\right\},x>0,\theta>0$$

且

$$\frac{\partial}{\partial\theta}\ln p(x;\theta)=-\frac{1}{\theta}+\frac{x}{\theta^2}=\frac{x-\theta}{\theta^2}$$

所以

$$I(\theta)=E\left(\frac{x-\theta}{\theta^2}\right)^2=\frac{\mathrm{Var}(x)}{\theta^4}=\frac{1}{\theta^2}$$

6.4.1 均方误差

问题：有偏估计一定是不好的估计吗?

- 无偏估计不一定比有偏估计更优.
- 评价一个点估计的好坏一般可以用：点估计值$\hat{\theta}$与参数真值θ的距离平方的期望, 这就是下式给出的**均方误差**(mean squared error)

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2$$

- 均方误差是评价点估计的最一般的标准. 我们希望估计的均方误差越小越好.

注意到

$$\begin{aligned}\mathrm{MSE}(\hat\theta) &= E(\hat\theta-\theta)^2 = E[(\hat\theta - E\hat\theta)+(E\hat\theta-\theta)]^2\\ &= E(\hat\theta-E\hat\theta)^2 + (E\hat\theta-\theta)^2 + 2E[(\hat\theta-E\hat\theta)(E\hat\theta-\theta)]\\ &= \mathrm{Var}(\hat\theta) + (E\hat\theta-\theta)^2\end{aligned}$$

均方误差由**点估计的方差与偏差的平方**两部分组成, 因此

- (1) 若$\hat\theta$是θ的无偏估计, 则$\mathrm{MSE}(\widehat{\theta}) = \mathrm{Var}(\hat\theta)$, 这说明这说明用$MSE$评价无偏估计和用点估计方差(有效性)评价无偏估计是完全一样的, 也说明用方差考察无偏估计有效性是合理的.
- (2) 当 $\hat\theta$不是θ的无偏估计时, 就要看其均方误差$\mathrm{MSE}(\hat\theta)$.
- 下面的例子说明: 在均方误差的含义下有些有偏估计优于无偏估计.

例6.4.1 对均匀总体$U(0,\theta)$, 由θ的极大似然估计$x_{(n)}$(见**例6.3.5**),但$x_{(n)}$不是无偏估计, 因为$E\left(X_{(n)}\right)=\frac{n}{n+1}\theta$(见**例6.2.5**).

修偏得到的无偏估计是 $\hat{\theta}=(n+1)x_{(n)}/n$ (见**例6.1.6**), 它的均方误差是

$$MSE(\hat{\theta})=\mathrm{Var}(\hat{\theta})=\frac{\theta^2}{n(n+2)}$$

现我们考虑θ的形如$\hat{\theta}_\alpha = \alpha \cdot x_{(n)}$的估计, 其均方误差为(计算过程见教材)

$$\mathrm{MSE}\left(\hat{\theta}_\alpha\right) = \alpha^2 \frac{n}{(n+1)^2(n+2)}\theta^2 + \left(\frac{n\alpha}{n+1} - 1\right)\theta^2$$

用求导的方法不难求出在$n \geq 2$时, 当$\alpha = (n+2)/(n+1)$ 时上述均方误差达到最小, 且其均方误差

$$\mathrm{MSE}\left(\frac{n+2}{n+1}x_{(n)}\right) = \frac{\theta^2}{(n+1)^2}$$

所以在均方误差的标准下, 有偏估计优于无偏估计.

在大样本时, 一般无需过多强调估计的有偏或无偏, 仅需保证一致. 拓展: 在高级计量中会学到, 为解决内生性为题, 尝试用工具变量估计方法(IV Estimation). IV估计是有偏的(但是是一致的).

作业3

A. 提交期末研究报告分组名单和拟选题

B. 6.3课后习题： 1, 2, 4, 7, 8, 10

贝叶斯估计略.

- 经典学派的观点: 统计推断是根据样本信息对总体分布或总体的特征数进行推断, 这里用到两种信息: 总体信息和样本信息.
- 贝叶斯学派的观点: 除了上述两种信息以外, 统计推断还应该使用第三种信息: 先验信息.

§6.6 区间估计(Interval estimation)

- 区间估计的概念
- 枢轴量法
- 单个正态总体参数的置信区间
- 两个正态总体下的置信区间
- 大样本情况的置信区间估计

6.6.1 区间估计的概念

想象你想经营一个食品商店. 问能否根据下面的市场调查结果进行决策?

(1) 点估计: 饮料的每日平均需求量是300瓶. 由矩估计原理, 300是期望需求量的点估计.

- 参数的点估计给出了具体数值, 但其精度如何, 点估计本身无法回答
- 区间估计给出未知参数的一个区间, 是度量一个点估计精度的方法

设θ是总体的一个参数, $x_1, ..., x_n$是样本, 区间估计就是要找两个统计量$\hat{\theta}_L = \hat{\theta}_L(x_1, \cdots, x_n)$和$\hat{\theta}_U = \hat{\theta}_U(x_1, \cdots, x_n)$, $(\hat{\theta}_L < \hat{\theta}_U)$, 使$\theta$以一定的概率(置信水平)落在$\left[\hat{\theta}_L, \hat{\theta}_U\right]$的区间.

注意:

- θ是参数, 不是随机变量.
- $\hat{\theta}_L, \hat{\theta}_U$是统计量, 所以都是随机变量.

由于样本的随机性, $\hat{\theta}_L, \hat{\theta}_U$也有随机性, 盖住未知参数$\theta$的可能性并不确定. 如果要对$\theta$的估计准确度有足够信心, 就要要求$P(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U)$尽可能大, 但这必然导致区间长度增大. 区间估计失去意义.

同样想象你经营一个食品商店. 调查公司给出的市场调研结果是

(2) 饮料的每日平均需求量是300瓶, 且有99.99%的概率每日需求量在0到600瓶间.

为解决此问题, 通常事先给定$\left[\hat{\theta}_L, \hat{\theta}_U\right]$盖住$\theta$的概率 (称为**置信区间**). 下面给出严格定义.

定义6.6.1 设θ是总体的一个参数, 其参数空间为Θ, $x_1, x_2, \ldots, x_n$是来自该总体的样本, 对给定的一个$\alpha(0 < \alpha < 1)$, 若有两个统计量 $\hat{\theta}_L = \hat{\theta}_L(x_1, \cdots, x_n)$ 和 $\hat{\theta}_U = \hat{\theta}_U(x_1, \cdots, x_n)$, 若对任意的$\theta \in \Theta$, 有

$$P_\theta\left(\hat{\theta}_L \leq \theta \leq \hat{\theta}_U\right) \geq 1 - \alpha$$

则称随机区间$\left[\hat{\theta}_L, \hat{\theta}_U\right]$为$\theta$的**置信水平(confidence level)**为$1-\alpha$**的置信区间(confidence interval)**, 或简称$\left[\hat{\theta}_L, \hat{\theta}_U\right]$是$\theta$ 的$1-\alpha$置信区间.

$\hat{\theta}_L$和 $\hat{\theta}_U$分别称为θ的(双侧)**置信下限(lower confidence limit)**和**置信上限**.

- 这里置信水平$1-\alpha$的含义是指在大量使用该置信区间时, 至少有$100(1-\alpha)\%$的区间含有θ.

例6.6.1 设$x_1, x_2, ..., x_{10}$是来自$N(\mu, \sigma^2)$的样本, 则μ的置信水平为$1-\alpha$的置信区间为

$$\left[\ \bar{x} - t_{1-\alpha/2}(9)s/\sqrt{10},\quad \bar{x} + t_{1-\alpha/2}(9)s/\sqrt{10}\ \right]$$

这里, $t_{1-\alpha/2}(9)$是自由度为9的t分布的$1-\alpha/2$分位数.

其中, $\bar{x}, s$分别为样本均值和样本标准差.

这个置信区间的由来将在6.6.3节中说明(σ^2未知用基于t分布的置信区间构造方法, default situation in the real world, so as in STATA), 这里用它来说明置信区间的含义.

若取$\alpha = 0.10$, 则$t_{0.95}(9) = 1.8331$. 上式化为

$$[\bar{x} - 0.5797s,\quad \bar{x} + 0.5797s]$$

现假定$\mu = 15$, $\sigma^2 = 4$, 则我们可以用随机模拟方法由$N(15, 4)$产生一个容量为10的样本, 如下即是这样一个样本:

14.85	13.01	13.50	14.93	16.97
13.80	17.95	13.37	16.29	12.38

由该样本可以算得

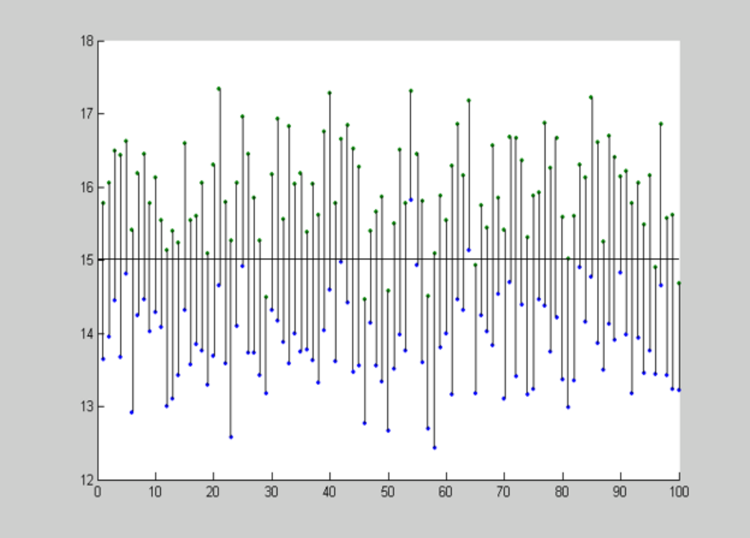
$$\bar{x} = 14.705, \quad s = 1.843$$

从而得到μ的一个区间估计为

$$14.705 \mp 0.5797 \times 1.843 = [13.637, 15.773]$$

该区间包含μ的真值15. 现重复这样的方法100次, 可以得到100个样本, 也就得到100个区间, 我们将这100个区间画在下图.

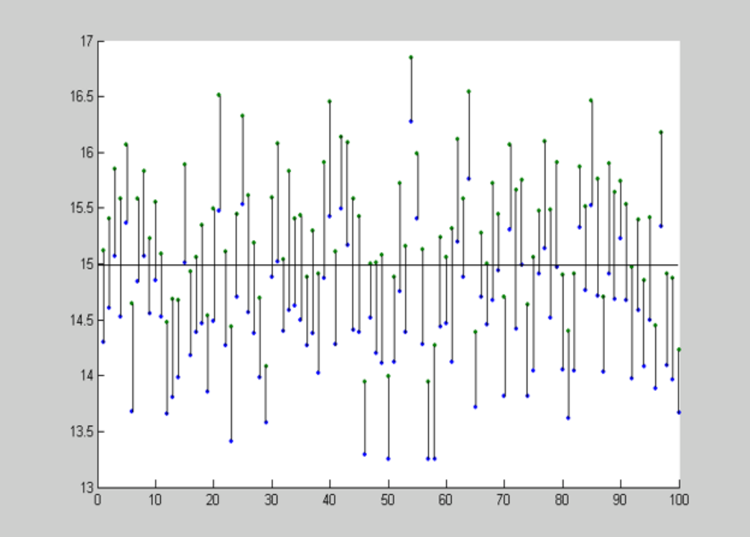
μ的置信水平为0.90的置信区间



这100个区间中有91个包含参数真值15，另外9个不包含参数真值.

μ的置信水平为0.5的置信区间

取$\alpha = 0.50$, 我们也可以给出100个这样的区间, 见下图. 这100个区间中有50个包含参数真值15.



上述模拟的stata code样例

```
clear

cap prog drop myprog
prog def myprog, rclass
 clear
 set obs 10
 gen x = rnormal(15, 2)
 ci mean x, level(90)
 ret scalar mean = r(mean)
 ret scalar ub = r(ub)
 ret scalar lb = r(lb)
end

simulate mean = 15 ub = r(ub) lb = r(lb), reps(100): myprog

gen x = _n
tw rspike ub lb x, yline(`=mean[1]')
```

Stata code：正态总体不同样本量下置信区间

```
//visualize confidence intervals (ci) of \mu of N(15, 2^2),
// at 90% confidence level and 50% confidence level

//generate 10 random numbers from N(15, 2^2)
clear
set obs 10
gen x=rnormal(15,2)
ci means x, level(90)
ci means x, level(50)

// generate 100 random numbers from N(15, 2^2)
clear
set obs 100
gen x=rnormal(15,2)
ci means x, level(90)
ci means x, level(50)

// What do you find from these results?

// How to calculate the ci of \sigma^2 of  N(15, 2^2)
help ci
ci variances x
ci variances x, level(90)
```

在总体为连续的场合, 可以用等式定义置信区间.

定义6.6.2 若对给定的$\alpha(0<\alpha<1)$, 和任意的$\theta\in\Theta$, 有

$$P_\theta\left(\hat{\theta}_L\leqslant\theta\leqslant\hat{\theta}_U\right)=1-\alpha$$

称$\left[\hat{\theta}_L,\hat{\theta}_U\right]$为$\theta$的$1-\alpha$**同等置信区间**. 在总体为连续分布场合下可以实现.

定义6.6.3, 6.6.4 若对给定的$\alpha(0<\alpha<1)$和任意的$\theta\in\Theta$, 有

$$P_\theta\left(\hat{\theta}_L\le\theta\right)\ge 1-\alpha$$

则称$\hat{\theta}_L$为θ的置信水平为$1-\alpha$的**(单侧)置信下限.** 假如等号对一切$\theta\in\Theta$成立, 则称$\hat{\theta}_L$为θ的$1-\alpha$同等置信下限.

若对给定的$\alpha(0<\alpha<1)$, 和任意的$\theta\in\Theta$, 有

$$P_\theta\left(\hat{\theta}_U\ge\theta\right)\ge 1-\alpha$$

则称$\hat{\theta}_U$为θ的置信水平为$1-\alpha$的**(单侧)置信上限.** 假如等号对一切$\theta\in\Theta$成立, 则称$\hat{\theta}_L$为θ的$1-\alpha$同等置信上限.

单侧置信上限和下限是置信区间的特殊情形. 因此, 寻求置信区间的方法可以用来寻找单侧置信限.

6.6.2 枢轴量法(Pivotal quantity)

下面用一个例子引出**用枢轴量法求区间估计**的方法.

例：已知正态总体方差σ^2, 求总体均值μ的置信区间.

给定**正态总体**: $X \sim N(\mu, \sigma^2)$, 总体方差σ^2已知. 在总体中抽取一个容量为n的样本$x_1, \ldots x_n$. 根据样本均值的性质, 有:

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

枢轴量G： 分布与未知参数μ**无关的**, 关于μ和样本的函数.

枢轴量法即要构建一个**与未知参数无关的**性状良好的**已知分布**, 以便运用该分布性质构造置信区间. (**思考:** 如正态总体方差σ^2未知, G能否继续做枢轴量?)

由标准正态分布性质, 对任意一个非负数c

$$P\left\{\left|\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right|\leq c\right\}=2\Phi(c)-1$$

即

$$P_{\mu}(\bar{x}-c\sigma/\sqrt{n}\leqslant\mu\leqslant\bar{x}+c\sigma/\sqrt{n})=2\Phi(c)-1=1-\alpha$$

所以$c=u_{1-\alpha/2}$ ($u_{1-\alpha/2}$为标准正态分布的$1-\alpha/2$分位数)

上式满足了置信区间的定义, 即以$1-\alpha$的概率保证μ落在内

$$\left[\ \bar{x}-u_{1-\alpha/2}\sigma/\sqrt{n},\ \ \bar{x}+u_{1-\alpha/2}\sigma/\sqrt{n}\ \right]$$

或μ的$1-\alpha$置信区间是

$$\left[\ \bar{x}-u_{1-\alpha/2}\sigma/\sqrt{n},\ \ \bar{x}+u_{1-\alpha/2}\sigma/\sqrt{n}\ \right]$$

6.6.2 枢轴量法

构造未知参数θ的置信区间的最常用的方法是枢轴量法, 其步骤可以概括为如下三步:

- 1. 设法构造一个样本和未知参数θ的函数 $G = G(x_1, x_2, ..., x_n, \theta)$ 使得G的分布不依赖于未知参数. 一般称具有这种性质的G为**枢轴量**(pivotal quantity).
- 2. 适当地选择两个常数c, d, 使对给定的 $\alpha(0 < \alpha < 1)$ 有

$$P(c \leq G \leq d) = 1 - \alpha$$

- 3. 假如能将$c \leq G \leq d$, 进行不等式等价变形化为$\hat{\theta}_L \leq \theta \leq \hat{\theta}_U$, 则$\left[\hat{\theta}_L, \hat{\theta}_U\right]$是$\theta$ 的$1-\alpha$**同等置信区间**.

关于置信区间的构造有两点说明:

- 满足置信度要求的c与d通常不唯一. 若有可能, 应选平均长度$E\left(\hat{\theta}_U - \hat{\theta}_L\right)$ 达到最短的c与d, 这在G的分布为对称分布场合通常容易实现.
- 实际中, 选平均长度$E\left(\hat{\theta}_U - \hat{\theta}_L\right)$尽可能短的$c$与$d$, 这往往很难实现. 因此, 常这样选择$c$与$d$, 使得两个尾部概率各为$\alpha/2$, 即$P(G < c) = P(G > d) = \alpha/2$, 这样的置信区间称为等尾置信区间. 这是在$G$的分布为偏态分布场合常采用的方法.

例6.6.2 设$x_1, x_2, \ldots, x_n$是来自均匀总体$U(0,\theta)$的一个样本, 试对给定的$\alpha(0<\alpha<1)$给出θ的$1-\alpha$ 同等置信区间.

解: (1) 前面已知θ的最大似然估计就是样本的最大次序统计量$x_{(n)}$, 其密度函数是

$$p(y) = ny^{n-1}/\theta^n, \quad y < \theta$$

根据随机变量函数密度函数的计算公式

$$p_Y(y) = \begin{cases} p_X[h(y)]h'(y), & a<y<b, \\ 0, & \text{其他}. \end{cases},$$

$x_{(n)}/\theta$的密度函数为$p(y;\theta) = ny^{n-1}, 0<y<1$, 与参数$\theta$无关, 故可取$x_{(n)}/\theta$作为枢轴量$G$.

(2) $x_{(n)}/\theta$的分布函数为$F(y) = y^n, 0<y<1$, 故

$$P\left(c \le x_{(n)}/\theta \le d\right) = d^n - c^n$$

因此我们可以适当地选择c和d满足$d^n - c^n = 1-\alpha$

(3) 利用不等式变形可给出θ的$1-\alpha$同等置信区间为$[x_{(n)}/d, x_{(n)}/c]$, 该区间的平均长度为

$$\left(\frac{1}{c}-\frac{1}{d}\right) Ex_{(n)}$$

这是因为,

$$P\left(c \le x_{(n)}/\theta \le d\right) = P\left(x_{(n)}/d \le \theta \le x_{(n)}/c\right) = d^n - c^n = 1-\alpha$$

可以求出, 在$0 \leqslant c < d \leqslant 1$及$d^n - c^n = 1-\alpha$的条件下, 当$d=1, c=\sqrt[n]{\alpha}$时, $\left(\frac{1}{c}-\frac{1}{d}\right)$ 取得最小值, 这说明$\left[x_{(n)}, x_{(n)}/\sqrt[n]{\alpha}\right]$是$\theta$的置信水平为$1-\alpha$的最短置信区间.

($0 \leqslant c < d \leqslant 1$是因为$\leqslant x_{(n)}/\theta$的支撑(support)是$(0,1)$.)

(推导思路: 首先证明d仅能取1, 因为$c=\sqrt[n]{d^n-1+\alpha}$只有$d=1$时能保证对所有n和α可以开方)

6.6.3 单个正态总体参数的置信区间

正态总体$N(\mu,\sigma^2)$是最常见的总体, 前文已知, μ可用$\bar{x}$估计(点估计), 其分布为$N(\mu,\sigma^2/n)$; σ^2可用样本方差s^2作为点估计.

下面针对正态总体详细讨论两个参数的置信区间.

一. σ已知时, μ的置信区间

在这种情况下, 枢轴量可选为

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

c和d应满足

$$P(c\leq G\leq d)=\Phi(d)-\Phi(c)=1-\alpha$$

经过不等式变形可得

$$P_\mu(\bar{x}-d\sigma/\sqrt{n}\leq\mu\leq\bar{x}-c\sigma/\sqrt{n})=1-\alpha$$

该区间长度为$(d-c)\sigma/\sqrt{n}$. 在$\Phi(d)-\Phi(c)=1-\alpha$的条件下, 由于标准正态分布Φ的密度函数是单峰对称的, 所以当$d=-c=u_{1-\alpha/2}$时, $d-c$达到最小 (思考: 为什么?), 由此给出的同等置信区间为

$$\left[\ \bar{x}-u_{1-\alpha/2}\sigma/\sqrt{n},\quad \bar{x}+u_{1-\alpha/2}\sigma/\sqrt{n}\ \right]$$

这是一个以$\bar{x}$ 为中心, 半径为$u_{1-\alpha/2}\sigma/\sqrt{n}$的对称区间, 常将之表示为

$$\bar{x}\pm u_{1-\alpha/2}\sigma/\sqrt{n}$$

课堂练习(例6.6.3)

例6.6.3 用天平秤某物体的重量9次, 得平均值为15.4 (克), 已知天平秤量结果为正态分布, 其标准差为0.1克. 试求该物体重量的0.95置信区间.

例6.6.3 用天平秤某物体的重量9次, 得平均值为15.4 (克), 已知天平秤量结果为正态分布, 其标准差为0.1克. 试求该物体重量的0.95置信区间.

解: 此处$1-\alpha=0.95$, $\alpha=0.05$, 查表知$u_{0.975}=1.96$, 于是该物体重量的0.95置信区间为

$$\bar{x}\pm u_{1-\alpha/2}\sigma/\sqrt{n}=15.4\pm 1.96\times 0.1/\sqrt{9}=15.4\pm 0.0653$$

从而该物体重量的0.95置信区间为$[15.3347,\quad 15.4653]$.

课堂练习(例6.6.4)

例6.6.4 设总体为正态分布$N(\mu, 1)$, 为得到μ的置信水平为0.95的置信区间长度不超过1.2, 样本容量应为多大?

例6.6.4 设总体为正态分布$N(\mu,1)$, 为得到μ的置信水平为0.95的置信区间长度不超过1.2, 样本容量应为多大?

解: 由题设条件知μ的0.95置信区间为

$$\left[\ \bar{x}-u_{1-\alpha/2}/\sqrt{n},\quad \bar{x}+u_{1-\alpha/2}/\sqrt{n}\ \right]$$

其区间长度为$2u_{1-\alpha/2}/\sqrt{n}$, 它仅依赖于样本容量n而与样本具体取值无关. 现要求

$$2u_{1-\alpha/2}/\sqrt{n}\leqslant 1.2$$

立即有$n\geqslant(2/1.2)^2u_{1-\alpha/2}^2$.

由于$1-\alpha=0.95$, 所以$u_{1-\alpha/2}=1.96$, 从而$n\geqslant(5/3)^2\times 1.96^2=10.67\approx 11$. 即样本容量至少为11时才能使得$\mu$的置信水平为0.95的置信区间长度不超过1.2.

小结

正态总体, σ已知时, μ的置信区间为

$$\left[\ \bar{x} - u_{1-\alpha/2}\sigma/\sqrt{n},\ \ \bar{x} + u_{1-\alpha/2}\sigma/\sqrt{n}\ \right]$$

基本结论:

- 置信区间的中心是样本均值
- 置信水平$1-\alpha$越高, $u_{1-\alpha/2}$越大, 则置信区间(C.I.)越长
- 样本量n越大, 则置信区间越短

在同样的置信度下, 置信区间越窄, 估计的精度越高.

人们希望(二者存在权衡(tradeoff)):

1. 置信区间越窄越好 (抽样误差越小)

2. 置信水平$1-\alpha$越高越好(关于误差估计的结论更可靠)

思考：置信水平越高越好么？

接着想象你经营一个食品商店. 调查公司给出的市场调研结果是

(2) 饮料的每日平均需求量是300瓶, 且有99.99%的概率每日需求量在0到600瓶间.

(3) 饮料的每日平均需求量是300瓶, 且有95%的概率每日需求量在300 ± 50瓶间.

哪一个更有助于帮助做决策?

为了避免置信区间过长带来的不足, 同时考虑置信水平不能太低, 一般可使用置信水平为$1-\alpha=0.95$ 的置信区间, 此时$u_{1-\alpha/2}=1.96$. (0.90, 0.99水平也常用)

注意：上述总体均值置信区间 C.I. 的构造受以下两个限制条件的约束.

- 必须是正态总体
- 总体方差已知

之后我们会学习, 由于大数定律, 当样本容量充分大时, 这两个条件就都不重要了.

二. 正态总体, σ未知时μ的置信区间

(由于σ未知,

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

不能作为枢轴量!)

这时可用t统计量, 由于**推论5.4.2**,

$$t = \frac{\sqrt{n}(\bar{x} - \mu)}{s} \sim t(n-1)$$

因此t可以用来作为枢轴量. 完全类似于上一小节, 可得到μ的$1-\alpha$置信区间为

$$\left[\bar{x} - t_{1-\alpha/2}(n-1)s/\sqrt{n}, \bar{x} + t_{1-\alpha/2}(n-1)s/\sqrt{n}\right]$$

此处$s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2$ 是σ^2的无偏估计.

例6.6.5 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12只轮胎试用, 测得它们的寿命(单位：万公里)如下：

$$4.68 \quad 4.85 \quad 4.32 \quad 4.85 \quad 4.61 \quad 5.02$$

$$5.20 \quad 4.60 \quad 4.58 \quad 4.72 \quad 4.38 \quad 4.70$$

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. 经计算有$\bar{x} = 4.7092, s^2 = 0.0615$.

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

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

在实际问题中，由于轮胎的寿命越长越好，因此可以只求平均寿命的置信下限，也即构造单边的置信下限. 由于

$$P\left(\frac{\sqrt{n}(\bar{x}-\mu)}{s}<t_{1-\alpha}(n-1)\right)=1-\alpha$$

由不等式变形可知μ的$1-\alpha$置信下限为

$$\bar{x}-t_{1-\alpha}(n-1)s/\sqrt{n}$$

将$t_{0.95}(11)=1.7959$代入计算可得平均寿命μ的0.95置信下限为4.5806(万公里).

三. 正态总体, μ未知, σ^2的置信区间

由**定理5.4.1**, 若总体为正态分布, 则

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

该分布不依赖未知参数μ, 可作为枢轴量.

取枢轴量 $G = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$, 由于$\chi^2$分布是偏态分布, 寻找平均长度最短区间很难实现, 一般都用等尾置信区间: 采用χ^2的两个分位数$\chi^2_{\alpha/2}(n-1)$和$\chi^2_{1-\alpha/2}(n-1)$, 在χ^2分布两侧各截面积为$\alpha/2$的部分, 使得

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

由此给出σ^2的$1-\alpha$置信区间为

$$\left[(n-1)s^2/\chi^2_{1-\alpha/2}(n-1), \quad (n-1)s^2/\chi^2_{\alpha/2}(n-1)\right]$$

例6.6.6 某厂生产的零件重量服从正态分布$N(\mu,\sigma^2)$, 现从该厂生产的零件中抽取9个, 测得其重量为(单位：克)

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

试求总体标准差σ的0.95置信区间.

解： σ^2的$1-\alpha$置信区间为

$$\left[(n-1)s^2/\chi^2_{1-\alpha/2}(n-1), \quad (n-1)s^2/\chi^2_{\alpha/2}(n-1)\right]$$

由数据可算得$s^2=0.0325$, $n=9$. 查表知$\chi^2_{0.025}(8)=2.1797$, $\chi^2_{0.975}(8)=17.5345$, 代入可得$\sigma^2$的$1-\alpha$置信区间为

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

从而σ的0.95置信区间为: $[0.1218, 0.3454]$ (简单开方, **为什么？**).

为什么可以简单开方？

因为

$$
\begin{aligned}
&P\left(\chi^2_{\alpha/2} \le \frac{(n-1)s^2}{\sigma^2} \le \chi^2_{1-\alpha/2}\right)\\
&= P\left((n-1)s^2/\chi^2_{1-\alpha/2}(n-1) \le \sigma^2 \le (n-1)s^2/\chi^2_{\alpha/2}(n-1)\right)\\
&= P\left(\sqrt{(n-1)s^2/\chi^2_{1-\alpha/2}(n-1)} \le \sigma \le \sqrt{(n-1)s^2/\chi^2_{\alpha/2}(n-1)}\right)
\end{aligned}
$$

课堂练习

培养一个新品种的苹果, 这种苹果除了口感好和颜色鲜艳以外, 另一个重要特征是单个重量差异不大. 为了评估新苹果, 她随机挑选了25个测试重量(单位：克), 其样本方差为$s^2 = 4.25$. 设苹果的重量服从正态分布, 试求σ^2的置信度为95%置信区间.

解:

课堂练习

培养一个新品种的苹果, 这种苹果除了口感好和颜色鲜艳以外, 另一个重要特征是单个重量差异不大. 为了评估新苹果, 她随机挑选了25个测试重量(单位：克), 其样本方差为$s^2 = 4.25$. 设苹果的重量服从正态分布, 试求σ^2的置信度为95%置信区间.

解:

$$P\left\{\frac{(n-1)s^2}{\chi^2_{1-0.025}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{0.025}}\right\} = 1 - 0.05$$

查表得:

$$\chi^2_{0.975}(24) = 39.4, \chi^2_{0.025}(24) = 12.4$$

$$\frac{(25-1)\times 4.25}{39.4} = 2.59, \frac{(25-1)\times 4.25}{12.4} = 8.23$$

所以σ^2得置信区间为$(2.59, 8.23)$.

6.6.4 大样本置信区间

- 非正态总体场合下, 寻找枢轴量及其分布比较困难. 如何办?
- 在样本容量充分大时, 可以用渐近分布来构造近似的置信区间. 一个典型的例子是关于比例p的置信区间.

例: 设$x_1, \ldots, x_n$是来自$b(1, p)$的样本, 求p的$1-\alpha$置信区间.

由中心极限定理, 有$\bar{x} \sim AN\left(p, \frac{p(1-p)}{n}\right)$. 所以

$$u = \frac{\bar{x} - p}{\sqrt{p(1-p)/n}} \sim AN(0, 1)$$

u可以作为近似枢轴量. 对给定α, 有

$$P\left(\left|\frac{\bar{x} - p}{\sqrt{p(1-p)/n}}\right| \le u_{1-\alpha/2}\right) \approx 1 - \alpha$$

6.6.4 大样本置信区间

通过变形, 可得到置信区间为

$$\left[\frac{1}{1+\frac{\lambda}{n}}\left(\bar{x}+\frac{\lambda}{2n}-\sqrt{\frac{\bar{x}(1-\bar{x})\lambda}{n}+\frac{\lambda^2}{4n^2}}\right), \frac{1}{1+\frac{\lambda}{n}}\left(\bar{x}+\frac{\lambda}{2n}+\sqrt{\frac{\bar{x}(1-\bar{x})\lambda}{n}+\frac{\lambda^2}{4n^2}}\right)\right]$$

其中记$\lambda = u_{1-\alpha/2}^2$, 实用中通常略去$\lambda/n$项(因为$n$很大), 于是可将置信区间近似为

$$\left[\bar{x}-u_{1-\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}, \quad \bar{x}+u_{1-\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right]$$

(注意: $\sqrt{\lambda/n}$并未略去, 因为是低次项.)

例6.6.7 对某事件A作120次观察, A发生36次. 试给出事件A发生概率p的0.95置信区间.

解： 此处$n = 120$, $\bar{x} = 36/120 = 0.3$ 而$u_{0.975} = 1.96$, 于是p的0.95(双侧)置信下限和上限分别为$0.218, 0.382$

所求的置信区间为$[0.218, 0.382]$

6.6.4 大样本置信区间: 重要拓展

事实上, 在大样本的情况下(如$n \geq 30$), 即使我们**不知道总体的分布, 利用中心极限定理也可以给出近似的置信区间**.

根据**定理5.3.3** 设$x_1, x_2, \ldots, x_n$是来自某个总体的样本, $\bar{x}$为样本均值. 若总体分布未知或不是正态分布, 但 $E(x) = \mu$, $Var(x) = \sigma^2$, 则n较大时的**渐近分布为**$\bar{x} \sim AN\left(\mu, \sigma^2/n\right)$, 即

$$\frac{\sqrt{n}\,(\bar{x} - \mu)}{\sigma} \sim AN(0, 1)$$

上述定理有一个推论: 即使我们不知道总体的方差σ^2, 但知道总体方差的相合估计s^2, 用s替换上式中的σ, 依然有

$$\frac{\sqrt{n}\,(\bar{x} - \mu)}{s} \sim AN(0, 1)$$

证明超过本书的内容, 需要渐近性质的知识. 基本思路见后.

利用

$$\frac{\sqrt{n}\,(\bar{x}-\mu)}{s} \sim AN(0,1)$$

我们可给出μ的近似$1-\alpha$置信区间:

$$\left[\bar{x}-u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}, \bar{x}+u_{1-\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right]$$

Slutsky's theorem: If $X_n \to X$ in distribution and $Y_n \to a$, a constant, in probability, then

(a) $Y_n X_n \to aX$ in distribution.
(b) $X_n + Y_n \to X + a$ in distribution.

Proof (Normal approximation with estimated variance)

Suppose that $\frac{\sqrt{n}(\bar{X}_n-\mu)}{\sigma} \to N(0,1)$

but the value σ is unknown. We know $s_n \to \sigma$ in probability ($plim(s_n) = \sigma$). $\sigma/s_n \to 1$ in probability (consistent estimator). Hence, Slutsky's theorem tells us

$$\frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{s_n} = \frac{\sigma}{s_n} \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma} \to N(0,1)$$

6.6.5 样本量的确定

样本量越大, 估计精度越高, 但大样本需要的成本高, 所以实际中人们关心, 在一定的精度要求下, 至少需要多大的样本量.

例6.6.8 某传媒公司欲调查电视台某综艺节目收视率p, 为使得 p 的$1-\alpha$置信区间长度不超过d_0, 问应调查多少用户?

解: 这是关于二点分布比例p的置信区间问题, 由两点分布p的$1-\alpha$置信区间

$$\left[\bar{x} - u_{1-\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}, \quad \bar{x} + u_{1-\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right]$$

知, $1-\alpha$的置信区间长度为$2u_{1-\alpha/2}\sqrt{\bar{x}(1-\bar{x})/n}$. 这是一个随机变量, 但由于$\bar{x}\in(0,1)$, 所以对任意的观测值有 $\bar{x}(1-\bar{x})\le 0.5^2 = 0.25$.

这也就是说p的$1-\alpha$的置信区间长度不会超过$u_{1-\alpha/2}/\sqrt{n}$.
现要求p的的置信区间长度不超过d_0, 只需要$u_{1-\alpha/2}/\sqrt{n} \le d_0$即可, 从而

$$n \ge \left(\frac{u_{1-\alpha/2}}{d_0}\right)^2$$

- 这是一类常见的寻求样本量的问题. 比如, 若取$d_0 = 0.04, \alpha = 0.05$, 则

$$n \ge \left(\frac{u_{0.975}}{0.04}\right)^2 = \left(\frac{1.96}{0.04}\right)^2 = 2401$$

- 这表明, 要使综艺节目收视率p的0.95置信区间的长度不超过0.04, 则需要对2401个用户作调查.

思考： 延续上面例题, 如果在置信水平不变的情况下, 你要使目前所得到的置信区间的长度减少一半, 样本量应增加到目前样本容量的多少倍?

如果保持置信区间的长度不变, 样本容量的增加会使什么发生变化?

课堂练习

如果对1000名随机抽样的居民调查对一项政策的支持程度，平均得到80%的支持度，则政策支持的点估计是多少？样本标准差是多少？95%置信区间是多少？

答：

课堂练习

如果对1000名随机抽样的居民调查对一项政策的支持程度，平均得到80%的支持度，则政策支持的点估计是多少？样本标准差是多少？95%置信区间是多少？

答： 解法一 (利用两点分布性质)： 点估计0.8,

样本标准差求法1: di sqrt((0.8*0.2)), 得出 0.4

95%置信区间求法1 di 0.8-1.96*sqrt((0.8*0.2)/1000) , 得出 .775

di 0.8+1.96*sqrt((0.8*0.2)/1000) , 得出 .824

所以 [.775, .825]

解法二 (利用大样本渐近分布)

样本标准差求法2:

```
clear
set obs 1000
gen var1=1
replace var1=0 if _n<=200
sum, de
```

得出标准差0.4

95%置信区间求法2:

di 0.8-1.96*0.4/sqrt(1000)，得出 .775

di 0.8+1.96*0.4/sqrt((0.8*0.2)/1000)，得出 .825

所以 [.775, .825]

6.6.6 两个正态总体下的置信区间

设$x_1, ..., x_m$是来自$N(\mu_1, \sigma_1^2)$的样本, $y_1, ..., y_n$是来自$N(\mu_2, \sigma_2^2)$的样本, 且两个样本**相互独立**.

$\bar{x}$与$\bar{y}$分别是它们的样本均值, $s_x^2 = \frac{1}{m-1}\sum_{i=1}^{m}(x_i - \bar{x})^2$和$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ 分别是它们的样本方差.

下面讨论两个均值差和两个方差比的置信区间

- $\mu_1 - \mu_2$的置信区间
- σ_1^2/σ_2^2的置信区间

一. $\mu_1 - \mu_2$的置信区间

1. σ_1^2和 σ_2^2已知时

由定理5.4.1及随机变量函数运算, 此时有

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

枢轴量可取

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

沿用前面方法, $\mu_1 - \mu_2$的$1 - \alpha$置信区间为

$$\bar{x} - \bar{y} \pm u_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$$

2. $\sigma_1^2 = \sigma_2^2 = \sigma^2$**未知时**

由定理5.4.1及随机变量函数运算, 此时有

$$\bar{x} - \bar{y} \sim N\left(\mu_1 - \mu_2, \left(\frac{1}{m} + \frac{1}{n}\right)\sigma^2\right)$$

$$\frac{(m-1)s_x^2 + (n-1)s_y^2}{\sigma^2} \sim \chi^2(m+n-2)$$

由于$\bar{x}, \bar{y}, s_x^2, s_y^2$相互独立, 故可构造服从$t(m+n-2)$分布的枢轴量

$$t = \sqrt{\frac{mn(m+n-2)}{m+n}} \frac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{\sqrt{(m-1)s_x^2 + (n-1)s_y^2}} \sim t(m+n-2)$$

沿用前面方法, 可推出$\mu_1 - \mu_2$的$1-\alpha$置信区间为

$$\bar{x} - \bar{y} \mp \sqrt{\frac{m+n}{mn}} s_w t_{1-\alpha/2}(m+n-2)$$

其中

$$s_w^2 = \frac{(m-1)s_x^2 + (n-1)s_y^2}{m+n-2}$$

3. $\sigma_2^2/\sigma_1^2 = c$已知时

构造方法与第二种情况相同, 可推出$\mu_1 - \mu_2$的$1-\alpha$置信区间为

$$\bar{x} - \bar{y} \pm \sqrt{\frac{mc+n}{mn}} s_w t_{1-\alpha/2}(m+n-2)$$

其中

$$s_w^2 = \frac{(m-1)s_x^2 + (n-1)s_y^2/c}{m+n-2}$$

4. 当m和n都很大时的近似置信区间(常用)

当m和n都很大时, 则无需任何有关σ_1^2, σ_2^2的信息. 由中央极限定理可知

$$\frac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{\sqrt{\frac{s_x^2}{m} + \frac{s_y^2}{n}}} \sim AN(0, 1)$$

由此可推出$\mu_1 - \mu_2$的$1 - \alpha$近似置信区间为

$$\left[\bar{x} - \bar{y} - u_{1-\alpha/2}\sqrt{\frac{s_x^2}{m} + \frac{s_y^2}{n}}, \quad \bar{x} - \bar{y} + u_{1-\alpha/2}\sqrt{\frac{s_x^2}{m} + \frac{s_y^2}{n}}\right]$$

5. 一般小样本情况下的近似置信区间 (略)

二. σ_1^2/σ_2^2的置信区间

由于

$$(m-1)s_x^2/\sigma_1^2 \sim \chi^2(m-1), (n-1)s_y^2/\sigma_2^2 \sim \chi^2(n-1)$$

且s_x^2与s_y^2相互独立, 故可仿照F变量构造如下枢轴量

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

对给定的$1-\alpha$, 由

$$P\left(F_{\alpha/2}(m-1, n-1) \le \frac{s_x^2}{s_y^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \le F_{1-\alpha/2}(m-1, n-1)\right) = 1-\alpha$$

经不等式变形即给出σ_1^2/σ_2^2的$1-\alpha$置信区间

$$\left[\frac{s_x^2}{s_y^2} \cdot \frac{1}{F_{1-\alpha/2}(m-1, n-1)}, \quad \frac{s_x^2}{s_y^2} \cdot \frac{1}{F_{\alpha/2}(m-1, n-1)}\right]$$

例6.6.10 某车间有两台自动机床加工一类套筒，假设套筒直径服从正态分布。现在从两个班次的产品中分别检查了5个和6个套筒，得其直径数据如下(单位：厘米)：

甲班：$5.06, 5.08, 5.03, 5.00, 5.07$

乙班：$4.98, 5.03, 4.97, 4.99, 5.02, 4.95$

试求两班加工套筒直径的方差比$\sigma_{甲}^2/\sigma_{乙}^2$的0.95置信区间

解： $1-\alpha = 0.95$

$$F_{0.025}(4,5) = \frac{1}{F_{0.975}(5,4)} = \frac{1}{9.36} = 0.1068$$
$$F_{0.975}(4,5) = 7.39$$

由数据算得$s_{甲}^2 = 0.00107$, $s_{乙}^2 = 0.00092$, 故$\sigma_{甲}^2/\sigma_{乙}^2$的0.95置信区间为 $[0.1574, 10.8899]$，没有理由说明甲机床的加工成品波动大。

置信区间的STATA运算命令

help ci

复习和思考

1.总体未知参数矩估计的思想方法是什么? 试写出0-1分布, 二项分布$b(m,p)$, 泊松分布$P(\lambda)$, 均匀分布$U(a,b)$, 正态分布$N(\mu,\sigma^2)$中有关参数的矩估计式.

2.极大似然估计的主要步骤是什么?

3.未知参数的估计量与估计值有什么区别?

5.估计量的三个基本评价标准是什么? 你能理解它们的含义吗?

6.求参数置信区间的一般方法是什么? 对正态总体, 试从有关的统计量自行导出几类参数的置信区间?

7.置信度的含义是什么? 置信度, 区间长度和样本容量的关系怎样?

作业4

6.6课后习题： 1-6, 8-12

上机实验2

正态总体$\mu = 15$, $\sigma^2 = 4$, 通过模拟, 总结并报告下述结果:

1. 固定样本量$n = 100$和$\alpha = 0.05$, 观察重复次数10, 20和40时置信区间包含真值$\mu = 15$的频率是否接近置信度$1 - \alpha = 0.95$
2. 设置$\alpha = 0.10$, 其他保持不变, 重复1, 观察模拟结果; 并观察与1中置信区间长度对比效果(随α的变化).
3. 将1中样本量变成$n = 200$, 其他不变, 重复1, 观察模拟结果; 并观察与1中置信区间长度对比效果(随n的变化).