第2章　参数估计

第2.1节　点估计

第2.2节　估计量的优良标准

第2.3节　区间估计

东北大学数学系

第2.1节　点估计

总体 X 的分布函数 F 含有未知的参数 θ，θ所有可能的取值范围称为“参数空间”，记为Θ。

从这个总体中抽取了一组样本 X_1，…，X_n，相应的样本观察值是 x_1，…，x_n。

应该如何估计出 θ 的具体数值？

点估计就是利用样本构造一个合理的统计量：$g(X_1, \ldots, X_n)$；用它的观察值 $g(x_1, \ldots, x_n)$ 去作为作为 θ 的估计值。

例2.1.1　政府或者企业希望了解人们的作息习惯。Gallup 公司做过一项调查， 56 % 的美国人说他们习惯早起， 44 % 的认为自己是“夜猫子”。

例2.1.2　丁同学在一个体重仪上称她的体重，假定这个体重仪没有系统误差，每次称量的结果是真实重量μ加上一个随机误差 ε_k 。一般认为 $\varepsilon_k \sim N(0, \sigma^2)$，因此 n 次称量的结果

$$X_k = \mu + \varepsilon_k \sim N(\mu, \sigma^2)$$

你可以用这组数据中的任何一个，或者样本均值，或者是样本中位数等，作为 μ 的估计值。

常用的点估计方法

矩估计：用样本的有关矩去作为总体有关矩的估计。即样本均值作为总体期望的估计；样本方差作为总体方差的估计；样本中位数(或众数) 作为总体中位数(或众数) 的估计等 。

极大似然估计：

所有情况中 “看起来最象” 的那个估计

2.1.1 矩估计

K.Pearson 的矩估计理论

假定总体 X 有 m 个未知参数 θ_1，…， θ_m，而有关的原点矩 $V_k = EX^k$ 存在，则应该有：

$$\begin{cases} V_1 = g_1(\theta_1, \ldots, \theta_m) \\ V_2 = g_2(\theta_1, \ldots, \theta_m) \\ \cdots\cdots \\ V_m = g_m(\theta_1, \ldots, \theta_m) \end{cases}$$

理论上求解方程组可以得到

$$\begin{cases} \theta_1 = h_1(V_1, \ldots, V_m) \\ \theta_2 = h_2(V_1, \ldots, V_m) \\ \cdots\cdots \\ \theta_m = h_m(V_1, \ldots, V_m) \end{cases}$$

假如用样本的 k 阶原点矩 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ 作为总体 k 阶矩 V_k 的估计，则可以得到总体未知参数$\theta_1, \ldots, \theta_m$ 的估计。

理论依据：大数律。矩估计基本上都是依概率或者几乎处处收敛到未知参数。

矩估计需要注意的几个问题

(1) 总体的参数不能表示成矩的函数时 (一般是总体矩不存在) ，就不能使用矩估计；

(2) 如果能够用低阶的矩估计，就不要用高阶矩；

(3) 按照矩估计的理论应该用样本的二阶中心矩来估计总体的方差，但是在实际应用中人们总是采用样本方差作为总体方差的估计。

矩估计的最大优点是简单实用，与总体分布形式没有关系。只要知道总体随机变量一些矩存在，就可以做相应的矩估计。

例2.1.3　设总体 $X \sim U(0,\theta)$，θ是未知参数，X_1，…，X_n是一组样本，求θ的矩估计。

解．总体的未知参数 θ 可以通过期望与方差表示：

总体期望：$\frac{\theta}{2}$；　总体方差：$\frac{\theta^2}{12}$。

因此根据矩估计的思想，可以得到两个矩估计：

$\hat{\theta} = 2\overline{X}$ 或者是 $\hat{\theta} = 2\sqrt{3}S$

习惯上我们采用第一个估计量

□

例2.1.4　X_1，…，X_n 是来自总体 $X \sim N(\mu, \sigma^2)$ 的一组简单随机样本，求 μ, σ^2 的矩估计。

解.　显然有：$V_1 = \mu$，$V_2 = \sigma^2 + \mu^2$；即

$$\mu = V_1, \quad \sigma^2 = V_2 - V_1^2 。$$

因此得到：

$$\hat{\mu} = \overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{k=1}^{n} X_k^2 - (\frac{1}{n}\sum_{k=1}^{n} X_k)^2$$

$$= \frac{1}{n}\sum_{k=1}^{n} (X_k - \overline{X})^2$$

□

几个常见分布的矩估计

二项分布 $B(N,p)$，N 已知	$\hat{p}=\dfrac{\overline{X}}{N}$
均匀分布 $U(a,b)$	$\overline{X}\pm\sqrt{3(n-1)/n}S$
泊松分布 $P(\lambda)$	$\hat{\lambda}=\overline{X}$
参数为 θ 的指数总体	$\hat{\theta}=1/\overline{X}$
正态总体 $N(\mu,\sigma^2)$	$\hat{\mu}=\overline{X}\quad\hat{\sigma}^2=\dfrac{n-1}{n}S^2$ $\hat{\sigma}=\sqrt{(n-1)/n}S$

例2.1.5　随机取 8 个零件，测得直径是 (*mm*)

74.001，74.005，74.003，74.001，

74.000，73.993，74.006，74.002。

设总体期望与方差存在，求它们的矩估计。

解 首先计算这组样本的样本均值和样本方差，

$$\bar{x} = 74.001375, \quad s^2 = 1.5696 \times 10^{-5}$$

根据矩估计的思想，虽然不知道这组样本来自什么样的总体，仍然可以给出总体期望和方差的矩估计：

$$\hat{\mu} = 74.001375, \quad \hat{\sigma}^2 = 1.3734 \times 10^{-5}$$

□

两个估计

1. 总体百分比的估计

采用样本中的百分比作为估计值。

总体分布被认为是一个两点分布，参数 p 或者说总体期望 p 就是总体的百分比。

因此如果希望了解总体中具有某种属性的个体的比例，只需要从总体中抽取部分样本，以样本中具有这种属性的比例 p_s 作为估计。

2. 序列号估计

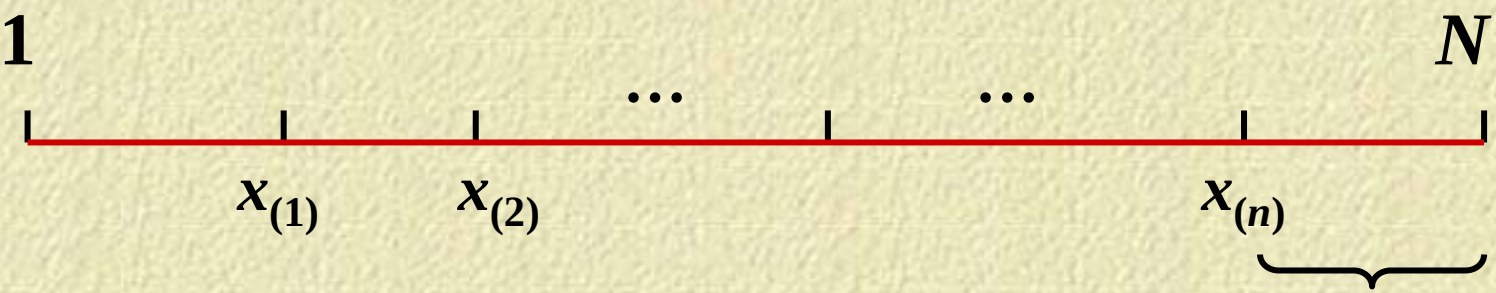
二战期间德国生产的每一辆坦克都带有一个编号(工厂出厂号)。盟军方面根据击毁或缴获的德军坦克的序列号：x_1，…，x_n，精确地估计出德国的坦克总产量 N。

解：

方法一. 假定每个得到的序列号 x_1，…，x_n 都是均匀地取自总体 {1，2，…，N}，那么这组样本数据居中的那一个数据(样本中位数)应该非常接近总体数据居中的那个数 (N/2)；

方法二. 仍然假定样本数据均匀取自总体数据，

则它们之间应该是等间隔地分布，因此总产量 N 就是样本中最大的那一个(极大统计量) 加上这些样本的 “平均间隔”。



这种估计：

$$x_{(n)} + \frac{x_{(n)} - x_{(1)}}{n-1} - 1$$

没有系统误差。

□

2.1.2 极大似然估计 (*MLE*)

1. 极大似然估计的想法

例2.1.6　假定盒子里黑、白球共 5 个，但是不知道黑球具体数目。现在随机有放回抽取 3 个小球，发现是两个黑球和一个白球。问盒子里最可能有几个黑球？

解：盒子里黑白球所有的可能有六种：

5白，4白1黑、3白2黑，2白3黑，1白4黑，5黑

以 p 记盒子里黑球所占的比例，

则 p 全部可能的值是：

$$\{\ 0,\ \frac{1}{5},\ \frac{2}{5},\ \frac{3}{5},\ \frac{4}{5},\ 1\ \}$$

定义三个统计量 X_1，X_2，X_3 表示抽样结果：取到黑球记为 1 ，否则记为 0 。因此 X_1，X_2，X_3独立同分布于参数 p 的两点分布。

例题中的三个样本观察值 x_1，x_2 ，x_3 有两个取值是 1，一个取值为 0。

而样本的联合分布律显然是

$$L(x,\ p) = p^{x1+x2+x3}(1-p)^{3-x1-x2-x3} = p^2(1-p)$$

注意这里样本的联合分布律

$$L(x,\ p) = p^{x1+x2+x3}(1 - p)^{3 - x1 - x2 - x3}$$

其实就是概率函数 $f(x, \theta)$ 。

它的含义是：当盒中黑球比例为 p 时，随机事件“有放回取出的三个小球中有两个黑球、一个白球”的概率。

对应于参数空间中不同的 p ，样本分布 $L(x,\ p) = p^2(1 - p)$ 所对应的这些概率是：

p	$0,\ \frac{1}{5},\ \frac{2}{5},\ \frac{3}{5},\ \frac{4}{5},\ 1$
$L(x, p)$	$0,\ \frac{4}{125},\ \frac{12}{125},\ \frac{18}{125},\ \frac{16}{125},\ 0$

既然“三个小球中包含两个黑球”是已经发生了的随机事件，因此使得这个事件发生概率取最大的那个值就是未知参数 p 最有可能的取值。

即 p 的极大似然估计就是 3/5。

□

R.A.Fisher 的极大似然估计理论

把概率函数 $f(x,\theta)$ 记为$L(x,\theta)$，并且认为

x 固定，它是 θ 的函数。

$L(x,\theta)$ 称为“似然函数”

1. 对离散总体，它是样本联合分布律；
2. 对连续总体，它是样本联合密度函数。

如果有 $L(x,\theta_1) < L(x,\theta_2)$，很自然我们会认为总体参数 θ 更有可能是 θ_2，而不太可能是 θ_1。

总体参数 θ 的极大似然估计就是使得似然函数在参数空间 Θ 中达到极大者

即对于任意 $\theta \in \Theta$ 都有：

$$\mathbf{L}(x,\hat{\theta}) = \max \mathbf{L}(x,\theta)$$

一般采用对数似然方程 (组) 求解*MLE*

$$\frac{\partial \ln \mathbf{L}(x,\theta)}{\partial \theta} = 0$$

无法建立似然方程时，必须根据定义求*MLE*

例2.1.7　设总体 $X \sim B(N, p)$，N 已知，p 是未知参数，X_1，…，X_n 是一组简单随机样本，求总体参数 p 的极大似然估计。

解.　不妨假定样本 X_1，…，X_n 相应的观察值是 x_1，…，x_n，而二项总体的似然函数为：

$$L(x,\theta) = [\prod \binom{N}{x_k}] p^{\sum x_k} (1-p)^{nN-\sum x_k}$$

这里每一个 $x_k = 0$、1、…、N 中的某个值

取对数再对参数 p 求导，得到对数似然方程：

$$\frac{\partial}{\partial\theta}\ln[L(x,\theta)]=\frac{\overline{x}}{p}-\frac{N-\overline{x}}{1-p}=0$$

因此，当 N 已知时，二项分布 $B(N,p)$ 中参数 p 的极大似然估计就是

$$\hat{p}=\frac{\overline{X}}{N}$$

两点分布的参数 p 的 MLE 就是样本均值

□

例2.1.8　X_1，…，X_n 是来自总体 $X \sim N(\mu, \sigma^2)$ 的简单随机样本，求 μ、σ^2 的极大似然估计。

解. 正态总体的似然函数为

$$L(x,\theta) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\{-\frac{1}{2\sigma^2}\sum_{k=1}^{n}(x_k-\mu)^2\}$$

注意这里总体参数 θ 是一个向量 (μ, σ^2)，因此对于似然函数取对数后分别对 μ, σ^2 求导，建立对数似然方程组：

$$\begin{cases} \dfrac{1}{\sigma^2}(\bar{x}-\mu)=0 \\ -\dfrac{n}{2\sigma^2}+\dfrac{1}{2(\sigma^2)^2}\sum_{k=1}^{n}(x_k-\mu)^2=0 \end{cases}$$

解方程组得到正态总体两个参数的*MLE*

$$\hat{\mu}=\overline{X} \qquad \hat{\sigma}^2=\frac{1}{n}\sum_{k=1}^{n}(X_k-\overline{X})^2=\frac{n-1}{n}S^2$$

□

练习**2.1.9**

总体标准差σ的极大似然估计是什么？如果μ已知，方差σ^2的极大似然估计又是什么？

例2.1.10　总体 $X \sim U(\theta, \theta+1)$，$\theta$ 是未知参数，X_1，…，X_n 是一组样本，求 θ 的极大似然估计。

解. 总体的密度函数为：

$$f(x, \theta) = 1, \quad \theta < x_1, \ldots, x_n < \theta+1$$

显然不能对参数 θ 求导，无法建立似然方程

注意到这个似然函数不是 0 就是 1，利用顺序统计量，把似然函数改写成如下形式：

$$f(x,\theta) = 1, \quad \theta < x_{(1)} < \ldots < x_{(n)} < \theta + 1$$

因此只要 $\theta < x_{(1)}$ 并且 $x_{(n)} < \theta + 1$ 同时满足，似然函数就可以达到极大值 1。

所以 $U(\theta, \theta+1)$ 中参数 θ 的极大似然估计可以是区间 $(x_{(n)} - 1, x_{(1)})$ 里的任意一个点。 □

说明 *MLE* 可以不唯一，甚至有无穷多个

同理，总体 $U(a, b)$ 左右端点 a、b 的 *MLE* 分别就是两个极值统计量 $x_{(1)}$、$x_{(n)}$。

几个常见分布的极大似然估计

二项分布 $B(N,p)$，N 已知	$\hat{p}=\dfrac{\overline{X}}{N}$
均匀分布 $U(a,b)$	$X_{(1)}$、$X_{(n)}$
泊松分布 $P(\lambda)$	$\hat{\lambda}=\overline{X}$
参数为 θ 的指数总体	$\hat{\theta}=1/\overline{X}$
正态总体 $N(\mu,\sigma^2)$	$\hat{\mu}=\overline{X}\quad \hat{\sigma}^2=\dfrac{n-1}{n}S^2$ $\hat{\sigma}=\sqrt{(n-1)/n}S$

2.1.3 极大似然估计与矩估计的简单比较

矩估计由 **K.Pearson** 在**1894**年提出，只要求总体的矩存在即可，不需要知道总体分布。

极大似然估计由 **R . A . Fisher** 在 **1912** 年提出的，必须要知道总体来自哪一种分布类型。

一般来说极大似然估计具有更多数学上的优良性，应用得更为广泛。

例如它肯定是充分统计量的函数；多数情况下在无偏估计的类中具有最小方差；具有渐近正态性；还是均匀先验分布时后验分布的概率函数的众数。

例2.1.11　为了研究密歇根湖湖滩地区的岩石成分，随机取了 100 个样品。每个样品中包含 10 个石子，记录下每个样品里属于石灰石的石子个数，有关数据为：

样品中石灰石子的个数	0	1	2	3	4	5	6	7	8	9	10
相应样品数	0	1	6	7	23	26	21	12	3	1	0

设 p 是这个地区的一块石子是石灰石的概率，则每个样品里的石灰石子的个数服从 $B(10, p)$，如果这 100 次观察是独立的，求 p 的估计。

解. 可以证明 p 的矩估计和极大似然估计都是：

$$\hat{p} = \frac{\overline{X}}{10}$$

关键的地方在于样本均值的计算，这里不是象过去那样，第一个样品有多少石灰石子、第二个样品有多少、…，一共 100 个数据全部一一罗列出来。而为了简便，给出的是包含多少个石灰石子的相应样品数。因此所有石灰石子个数有：$0\times0 + 1\times1 + 2\times6 + \ldots + 9\times1 + 10\times0 = 499$，一共观察了 100 次。

$$\overline{x} = \frac{499}{100}, \quad \hat{p} = 0.499$$

□

练习2.1.12

如果没有砝码，我们应该如何去判断一个测量仪器的精度(即标准差)？

练习2.1.13

$X_1, \ldots, X_n$ 是来自负二项分布的一组样本，

$$p_k = C_{k-1}^{r-1} p^r q^{k-r}, \quad k \geq r$$

求总体参数 p、r 的矩估计以及极大似然估计。

练习2.1.14

$X_1, \ldots, X_n$ 是来自*Logistic*分布的一组样本，

$$F(x) = \frac{1}{1+e^{-\alpha x-\beta}}, \quad \alpha > 0$$

求总体参数α、β 的极大似然估计。

练习2.1.15

$X_1, \ldots, X_n$ 是来自对数正态分布的一组样本，

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x}\exp[-\frac{(\ln x-\mu)^2}{2\sigma^2}]$$

求总体参数 μ, σ^2 的矩估计和极大似然估计。

练习2.1.16

$X_1, \ldots, X_n$ 是来自*Weibull* 分布的一组样本，

$$f(x) = \mu\alpha x^{\alpha-1} \exp[-\mu x^{\alpha}], \quad x > 0$$

求总体参数 $\mu\,(>0), \alpha\,(>0)$ 的极大似然估计。

练习2.1.17

$X_1, \ldots, X_n$ 是来自*Pareto* 分布的一组样本，

$$f(x) = \alpha M^{\alpha} x^{-(1+\alpha)}, \quad x \geq M$$

求总体参数 $\alpha\,(>0), M\,(>0)$ 的极大似然估计。

第2.2节 估计的优良标准

一般的，一个良好的点估计应该满足三个标准：

无偏性：估计量的数学期望要等于参数；

有效性：估计量的方差要比较小(主要是限制在无偏估计的范围内)；

一致性：当样本容量趋于无限多时，估计量应该收敛到参数。

2.2.1 无偏估计(*Unbiased estimation*)

定义2.2.1 参数 $g(\theta)$ 的估计量 $\varphi(X_1,..,X_n)$ 如果满足：$E\,\varphi(X_1,..,X_n) = g(\theta)$ 对 Θ 中所有的 θ 都成立，则称 $\varphi(X_1,..,X_n)$ 是 $g(\theta)$ 的一个无偏估计量。

Remark

无偏性是估计好坏的一个基本要求，它表明即使每一次估计都可能有误差，但是从长远来看，这种估计总的误差能够相互抵消。

例2.2.1 假定总体 X 的期望 μ，方差 σ^2 存在，则样本均值、样本方差分别是 μ、σ^2 的无偏估计。

证明. 样本均值是 μ 的无偏估计很显然；

只需证明 $ES^2 = \sigma^2$。

$$S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2 = \frac{1}{n-1}\{\sum_{k=1}^{n} X_k^2 - n\overline{X}^2\}$$

$$E\,S^2 = \frac{1}{n-1}\{n[\mu^2+\sigma^2] - n[(E\overline{X})^2 + D\overline{X}]\}$$

$$= \frac{1}{n-1}\{n[\mu^2+\sigma^2] - n[\mu^2 + \frac{\sigma^2}{n}]\} = \sigma^2$$

□

练习2.2.2

对于总体 $X \sim N(\mu, \sigma^2)$，验证 $n \geq 2$ 时样本绝对偏差及样本标准差的统计量都是 σ 的无偏估计。

$$\varphi_1 = \sqrt{\frac{\pi}{2}} \frac{1}{\sqrt{n(n-1)}} \sum_{k=1}^{n} |X_k - \overline{X}|$$

$$\varphi_2 = \frac{\Gamma(\frac{n-1}{2})}{\sqrt{2}\Gamma(\frac{n}{2})} \sqrt{\sum_{k=1}^{n} (X_k - \overline{X})^2}$$

提示：考虑$N(0,1)$的绝对矩以及卡方分布的期望

例2.2.3　总体 $X \sim U(0,\theta)$，θ 是未知参数，讨论 θ 的无偏估计。

解. θ 的矩估计是 $2\overline{X}$，显然是无偏估计，不过样本均值不是充分统计量，因此可能不如极大似然估计 $X_{(n)}$ 好，但是 $X_{(n)}$ 是否无偏的？

$X_{(n)}$ 的分布函数显然是

$\mathbf{P}\{X_{(n)} \le x\} = (x/\theta)^n$，　$0 < x < \theta$。因此

$E X_{(n)} = \dfrac{n}{n+1}\theta$；修正后得到根据充分统计量构成的 θ 的 UE 是 $\dfrac{n+1}{n} X_{(n)}$。

□

利用充分统计量构造无偏估计

假定样本是 X_1，…，X_n，充分统计量为 T，参数 $g(\theta)$ 的无偏估计量是φ，则$E(\varphi|T)$ 是 $g(\theta)$ 的由充分统计量构成的无偏估计。

练习2.2.4

样本 X_1，…，X_n 来自参数 p 的两点分布，利用充分统计量 $T = X_1+\ldots+X_n$ 构造总体方差 $p(1-p)$ 的一个无偏估计量。

糟糕的无偏估计

反例1

设总体 X 来自泊松分布 $P(\lambda)$，现在只有一个样本 X_1，求 $g(\lambda) = e^{-2\lambda}$ 的无偏估计。

这里无偏估计只有一个：

当 X_1 的观察值为偶数时，用 1 估计 $e^{-2\lambda}$ ；
当 X_1 的观察值为奇数时，用 - 1 估计 $e^{-2\lambda}$

反例2

总体 X 来自两点分布 $B(1,p)$，仍然只有一个样本 X_1，则 p^2 的无偏估计不存在。

如果 $g(X_1)$ 是 p^2 的一个无偏估计，则多项式

$$p\,g(1) + (1-p)\,g(0) = p^2$$

对所有的 $0 < p < 1$ 都成立，矛盾。

所以这里无偏估计不存在。

2.2.2 有效性

1. 如何衡量估计的偏差

$$\varphi(X_1, \cdots, X_n) + Y$$

定义2.2.2 假定 $\varphi(X_1,..,X_n)$ 是 $g(\theta)$ 的一个估计，则

$$M(\varphi) = E[\varphi(X_1,..,X_n) - g(\theta)]^2$$

称为是估计量 $\varphi(X_1,..,X_n)$ 的均方误差(MSE)。

MSE 越小估计就越好，

UE 的 ***MSE*** 就是它的方差

例2.2.5　总体 $X \sim U(0,\theta)$，比较 θ 的两个无偏估计：$\varphi_1 = 2\overline{X}$ 与 $\varphi_2 = \frac{n+1}{n} X_{(n)}$ 的 MSE。

解. 由于都是无偏估计，因此只需计算方差

显然　$Var\ \varphi_1 = 4\ Var\ (X_1) / n = \frac{\theta^2}{3n}$，

容易计算出　$Var\ \varphi_2 = \frac{\theta^2}{n(n+2)}$；

$n = 1$时即只有一个样本，它们的 MSE 相同，但事实上这时这两个估计重合；

当样本容量大于1总有 $n(n+2) > 3n$，所以从均方误差的角度看，φ_2 也要比 φ_1 好。

□

2. 限制在UE 中的最优估计

定义2.2.3　一致最小方差无偏估计 ($UMVUE$)

假定 $\varphi_0(X_1,..,X_n)$ 是 $g(\theta)$ 的一个无偏估计，并且 $M(\varphi_0) \le M(\varphi)$ 对$g(\theta)$ 的任意 UE φ 都成立，则称$\varphi_0(X_1,..,X_n)$是$g(\theta)$ 的一致最小方差无偏估计

显然 $g(\theta)$ 的无偏估计的方差越小越好，但是这些方差不可能任意地小。 $g(\theta)$ 的所有无偏估计的方差有一个公共的下界(C-R下界)。

方差达到这个下界的 UE 自然就是 $UMVUE$

3. 一般情况下如何寻找 *UMVUE*

Blackwell-Lehmann-Sheffe 定理

如果T是充分、完备的统计量，$\varphi(T)$是$g(\theta)$的一个无偏估计，则$\varphi(T)$就是$g(\theta)$的*UMVUE*。

只需利用充分、完备统计量去构造无偏估计

思考1

给出计算任意$g(\theta)$的*UMVUE*的思路。

例2.2.6　求 $N(\mu, \sigma^2)$ 中参数 μ, σ^2 的*UMVUE*

解. 根据正态分布密度函数或因子分解定理，充分完备统计量是：

$$(\sum_{k=1}^{n} X_k, \sum_{k=1}^{n} X_k^2)$$

现在已知样本均值、样本方差分别是 μ, σ^2 的无偏估计，最重要的，它们还都是充分完备统计量的函数，因此 μ, σ^2 的*UMVUE* 分别就是样本均值和样本方差。

□

例2.2.7　关于一些常见分布的参数的*UMVUE*

N 已知时 $B(N,p)$，	$\hat{p}=\dfrac{\overline{X}}{N}$
泊松分布 $P(\lambda)$	$\hat{\lambda}=\overline{X}$
参数为 θ 的指数总体	$\hat{\theta}=1/\overline{X}$
$N(\mu,\sigma^2)$	$\hat{\mu}=\overline{X}\quad \hat{\sigma}^2=S^2$

这些估计量都是无偏估计，

a . 由因子分解定理，它们都是充分统计量；

b. 总体属于指数族，它们也都是完备统计量；

因此根据**B-L-S** 定理，这些估计都是*UMVUE*。

2.2.3 一致估计(*Consistent estimation*)

参数 $g(\theta)$ 的估计量 $\varphi(X_1,..,X_n)$ 总是与样本容量 n 有关，因此不妨记为 φ_n ；

一个好的估计直观上应该满足：

当 n 充分大时， φ_n 要充分接近 $g(\theta)$ 。

定义2.2.4 对于任意 $\varepsilon > 0$ ，如果当 $n \to \infty$ 时，

$$P\{|\varphi_n - g(\theta)| > \varepsilon\} \to 0$$

则称 φ_n 是 $g(\theta)$ 的一致估计(又叫相合估计)

即， φ_n 依概率收敛到 $g(\theta)$

强相合估计：

$$P\{\varphi_n \to g(\theta)\} = 1$$

即 φ_n不收敛到 $g(\theta)$ 的那些样本点的概率为 0

一般来说矩估计都具有强相合性，而在较广泛的条件下，极大似然估计也具有强相合性。

渐近正态估计：*Asymptotically normal estimation*

如果存在一个常数 $\sigma > 0$ ，使得

$$\frac{n^{1/2}[\varphi_n - g(\theta)]}{\sigma} \to N(0,1)$$

即， φ_n的分布可以近似认为是 $N(g(\theta), \frac{\sigma^2}{n})$

练习2.2.8

样本 $X_1, \ldots, X_n$ 来自泊松总体 $P(\lambda)$，构造参数λ 的渐近正态估计。

练习2.2.9

样本 $X_1, \ldots, X_n$ 来自总体 $U(0, \theta)$，已知$X_{(n)}$ 是总体参数θ的极大似然估计。证明它也是θ的一致估计。(甚至是强相合及任意阶的矩相合估计)

第2.3节 区间估计

矩估计与极大似然估计，都是一种点估计。

区间估计是指用一个(随机) 区间去做未知参数 $g(\theta)$ 的估计，这个区间称为是置信区间 。

这个区间包含 $g(\theta)$的概率称为置信度或置信水平；

区间的长度称为是这个区间估计的精度，

长度越短，即精度越高，这个区间越好。

区间估计的想法是“给所做的结论留些余地”，表示我们有多大的把握肯定我们所做的结论。

显然对于总体的未知参数一个区间要比一个数值提供的信息更多，也更让人放心。

置信度越大，则区间的长度应该越长，即精度小，或者说抽样误差大。

在实际的统计应用中大多数的置信区间是由样本统计量±“抽样误差”来构造。

2.3.1 置信区间理论(*Confidence interval*)

定义2.3.1 给定一个常数 $0 < \alpha < 1$，对于总体未知参数 $g(\theta)$，如果存在两个统计量 φ_1、φ_2 满足：

$$P\{\varphi_1(X) < g(\theta) < \varphi_2(X)\} \geq 1 - \alpha$$

则称(φ_1, φ_2)是$g(\theta)$的置信度$1 - \alpha$的置信区间；

φ_1、φ_2分别被称为是置信下限与置信上限。

有时也只考虑单侧区间 $(\varphi_1, +\infty)$ 或 $(-\infty, \varphi_2)$

点估计可以形式上认为是一种特殊的区间估计，这个区间的长度为 0 。

区间估计的置信度与精度是一对矛盾。

如果置信度越高，明显地区间应该越大，即误差大，区间的精度低。反之同理。

J . Neyman 的观点：

先考虑置信度，再去讨论估计的精度

先找出一些以1 - α 概率包含未知参数的区间，再从这些区间里去找长度最短者。

2.3.2 区间估计的求解思路

置信区间主要依据统计量的抽样分布或者是大样本理论来构造。

第一步　找一个枢轴变量 $Z(X,\theta)$。

枢轴变量是一个随机变量，它与抽取出的样本以及待估计的 $g(\theta)$ 都有关系。但是它的分布又必须是与参数 θ 无关的已知分布。

一般是从 $g(\theta)$ 的良好的点估计出发，去寻找枢轴变量 $Z(X,\theta)$。

第二步　对于给定的置信度 $1-\alpha$，求出两个常数 a、b，使得：

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

第三步　变换不等式，成为等价的形式：

$$a < Z(X,\theta) < b$$

$$\longleftrightarrow \varphi_1(X) < g(\theta) < \varphi_2(X)$$

因此区间 (φ_1, φ_2) 就是 $g(\theta)$ 的一个置信度为 $1-\alpha$ 的区间估计。

2.3.3 常见的几个区间估计

1. 总体属性比例的置信区间

假定从总体中抽取了n个观察值，以p_s记样本里具有某种属性的比例，则总体中具有这种属性的比例p的$1-\alpha$区间估计近似是：

$$p_s - u_{\alpha/2}\sqrt{\frac{p_s(1-p_s)}{n}} \quad \text{到} \quad p_s + u_{\alpha/2}\sqrt{\frac{p_s(1-p_s)}{n}}$$

$u_{\alpha/2}$ 恰好是标准正态分布的双侧 α 分位点

依据是 De Moivre – Laplace 中心极限定理

a. 随机从总体中抽取一个样本，它具有这种属性的概率是 p ；

b. 随机从大总体中抽取 n 个样本，其中具有这种属性的样本个数 X 近似有 $X \sim B(n, p)$ ；

c. 根据中心极限定理，又近似有：

$$\frac{X - np}{\sqrt{np(1-p)}} \sim N(0,1)$$

把上式改写成：

$$\frac{\frac{X}{n}-p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$$

这里(X/n) 正好是样本中的比例 p_s ，而根号符号里的 p 未知，因此用 p_s 近似替代，注意到标准正态分布是一个对称的分布，因此得到总体比例 p 的最短的近似区间估计：

$$p_s - u_{\alpha/2}\sqrt{\frac{p_s(1-p_s)}{n}} \quad \text{到} \quad p_s + u_{\alpha/2}\sqrt{\frac{p_s(1-p_s)}{n}}$$

例2.3.1　考虑容量为 1200 的一组样本构造的区间估计，假定有60% 的同学每天自习时间超过 2 小时。一个 95% 的置信区间的抽样误差是 2.8 个百分点。

这个结论的含义是：

有 95% 的把握可以肯定，学习认真同学的真实比例界于 57.2 ~ 62.8 之间。

同理，90% 的置信区间的抽样误差是2.3 个百分点，因此以 90% 的把握可以确定的这个区间是57.7 ~ 62.3 。

□

例2.3.2　随机询问500 名工作一年的大学毕业生，其中290 人表示对自己的工作还算满意，即新毕业大学生中有58%不反感自己的工作。

从区间估计的角度，95%区间的抽样误差是

$$1.96\sqrt{\frac{0.58(1-0.58)}{500}} = 2.2\%$$

有 55.8 ~ 60.2 的毕业生认可他的新工作。 □

练习2.3.3

不满意而想换工作的比例大约在什么范围？

2. 两个属性比例之差的置信区间

假定从两个总体中分别抽取了 n_1、n_2 个观察值，样本比例相应是 p_{s1}、p_{s2} 。则两个总体比例之差 $p_1 - p_2$ 的置信水平 $1 - \alpha$ 的区间估计近似地是：

$$(p_{s1} - p_{s2}) - u_{\alpha/2}\sqrt{\frac{p_{s1}(1-p_{s1})}{n_1} + \frac{p_{s2}(1-p_{s2})}{n_2}}$$

到

$$(p_{s1} - p_{s2}) + u_{\alpha/2}\sqrt{\frac{p_{s1}(1-p_{s1})}{n_1} + \frac{p_{s2}(1-p_{s2})}{n_2}}$$

例2.3.4 *Time / CNN* 曾经进行了一项电话委托调查：访问 503 名非洲裔美国人，询问他们是更喜欢用“非洲裔美国人” 还是“黑人”来作为他们种族的称呼，结果其中有26% 的人更喜欢第一种称呼；五年以后这个调查被重新做了一遍，发现有53% 更喜欢 “非洲裔美国人”的称呼。调查是否表明与五年前相比人们的观点有改变？

分析： 问题的关键是，这 27 个百分点的差异究竟是来自于调查时不可避免的随机误差，还是因为这两次调查时真实的比例的确发生了 “ 显著的 ” 改变？

首先计算抽样误差，

$$1.96\sqrt{\frac{0.26(1-0.26)}{503}+\frac{0.53(1-0.53)}{503}}=0.058$$

即真实比例的百分比差异 p_1 - p_2 以95%可能界于 27-5.8 = 21.2 到 27＋5.8 = 32.8 之间。

很显然这个区间(21.2，32.8) 不包含 0 点，因此可以认为真正的差异发生了变化。

进一步还可以认为大约有21% 到33%的比例变得更喜欢用“非洲裔美国人”来称呼他们自己。

□

3. 正态总体均值的置信区间

假定样本 X_1，..，X_n 来自总体 $N(\mu, \sigma^2)$

(1) 如果总体方差已知 $\sigma^2 = \sigma_0^2$

此时样本均值的分布为 $N(\mu, \frac{\sigma_0^2}{n})$

$$P\{|\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma_0}| \le u_{\alpha/2}\} = 1-\alpha$$

总体均值 μ 的
$1-\alpha$ 的区间估计：$(\overline{X} - u_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}, \overline{X} + u_{\alpha/2}\frac{\sigma_0}{\sqrt{n}})$

(2) 如果总体方差σ^2未知

需要使用抽样分布中定理1.3.1

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

因此有：

$$\mathrm{P}\{\,|\frac{\sqrt{n}(\overline{X}-\mu)}{S}|\le t_{\alpha/2}(n-1)\,\} = 1-\alpha$$

总体均值μ的1 - α区间估计为：

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

例2.3.5　科学上的很多重大发现往往由年轻人提出，下表是 16 世纪中期到 20 世纪的 12 项重大科学突破的情况：

科学发现	科学家	时间	年龄
日心说	哥白尼	1543	40
望远镜,天文学基本定律	伽利略	1600	43
动力学,万有引力,微积分	牛顿	1665	23
电的本质	富兰克林	1746	40
燃烧即氧化	拉瓦锡	1774	31

地球的演变	莱尔	1830	33
进化论	达尔文	1858	49
光的电磁特性	麦克思韦	1864	33
放射性	居里	1896	34
量子力学	普朗克	1901	43
狭义相对论	爱因斯坦	1905	26
量子力学的数学基础	薛定谔	1926	39

假定数据来自期望、方差未知时的正态总体，问什么年龄段科学家们将可能做出重要的工作？

解. 首先计算样本统计量，

$$\bar{x} \approx 36.17, \quad s \approx 7.53$$

现在有12个样本，因此抽样误差是

$$t_{0.025}(11)\frac{s}{\sqrt{12}} = 2.201 \times \frac{7.53}{3.4641} = 4.78$$

可以构造出一个区间 $(31.4, 41.0)$ □

历史数据表明，科学家研究工作的黄金时期是31岁半到41岁间。这个年龄段他们将有可能做出杰出的工作。

这个结论的可靠程度是 95% 。

4. 正态总体方差的置信区间

只讨论μ未知的情况，由抽样分布定理1.3.1

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

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

虽然卡方分布的密度函数不是对称函数，习惯上仍然取总体方差 σ^2的 1 - α 的区间估计为：

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

5. 两个正态总体均值差的置信区间

假定从总体 $X \sim N(\mu_1, \sigma_1^2)$ 中抽取 n_1 个样本，从另个独立的总体 $Y \sim N(\mu_2, \sigma_2^2)$ 中抽取 n_2 个样本；

相应的样本均值与样本方差分别为：

$$\overline{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i \quad , \quad S_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i - \overline{X})^2$$

$$\overline{Y} = \frac{1}{n_2}\sum_{j=1}^{n_2} Y_j \quad , \quad S_2^2 = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2}(Y_j - \overline{Y})^2$$

根据抽样分布中定理1.3.2，有

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

这里 $S_W^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$

因此均值差 μ_1 - μ_2 的1 - α 区间估计为：

$$(\overline{X}-\overline{Y}-t_{\alpha/2}(n_1+n_2-2)S_W\sqrt{\frac{1}{n_1}+\frac{1}{n_2}},$$

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

6. 两个正态总体方差比的置信区间

仍然根据抽样分布中定理1.3.2，有

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

因此方差比 σ_1^2/σ_2^2 的1 - α 区间估计为：

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

例2.3.6 用某种标准给子女和父母亲近程度打分，

样本容量	平均接触程度	标准差
71(父亲酗酒)	78分	25
46(父亲不酗酒)	91分	22

给出 95% 的区间估计并且讨论。

解. 对于父亲酗酒的子女构成的总体，

抽样误差是

$$t_{0.025}(70)\frac{25}{\sqrt{71}} \approx 5.9$$

一个95%的区间估计是 $(72.1, 83.9)$；

同理不酗酒父亲的子女构成总体的抽样误差是

$$t_{0.025}(45)\frac{22}{\sqrt{46}} \approx 6.6$$

一个95%的区间估计是 $(84.4, 97.6)$；

而这两个均值差的区间估计是 ($s \approx 23.9$)

$(4.1, 21.9)$

可以认为正常家庭子女比酗酒父亲的子女要更亲近父母，大约高出4～22分。

□

置信水平的理解

由于总体参数是未知的，对于每一个计算出来的区间估计，它要么包含总体参数，要么不包含，只是我们不知道而已 ；

但我们可以肯定，如果采用某种方法构造出一个置信水平 **0.95** 的区间(这个区间的两个端点是统计量的函数)，当我们代入 **100** 次统计量的数据从而得到**100** 个区间时，平均有 **95** 个区间要包含总体参数。

样本容量对区间长度的影响

以95%的区间估计为例，

总体比例 $2\times 1.96\sqrt{\dfrac{p_s(1-p_s)}{n}}$

两个比例之差 $2\times 1.96\sqrt{\dfrac{p_{s1}(1-p_{s1})}{n_1}+\dfrac{p_{s2}(1-p_{s2})}{n_2}}$

方差未知正态总体 $2\times t_{0.025}(n-1)\dfrac{s}{\sqrt{n}}$

方差已知正态总体 $2\times 1.96\dfrac{\sigma_0}{\sqrt{n}}$

4 倍的样本容量，抽样误差才可能缩减一半

民意调查中的估计

在总体比例的区间估计中抽样误差是：$1.96\sqrt{\frac{p_s(1-p_s)}{n}}$

根据不等式关系 $p_s(1-p_s) \le 0.5\times 0.5 = 0.25$，所以只要取样本容量 $n \ge 1200$，就足够保证抽样误差 ≤ 3 个百分点。

这也是大多数的民意调查至少要保证抽取1200个样本的原因，

同时抽样误差被近似成 $100/\sqrt{n}$ 个百分点。

2.3.4 序贯区间估计

一个好的区间估计，应该是置信度不小于事先给定的 $1-\alpha$，而区间的长度又不要超过事先指定的某个常数 l。

这种区间是否存在？又如何去构造？

序贯或逐次 (*Sequential*) 方法最初是 **A. Wald** 在二战后期为处理美国军火生产中质量检验问题而提出的一个一个地抽取样本的方法 (序贯概率比检验) :　样本容量 n 不再事先给定，而是根据抽样或观测过程来决定什么时候停止抽样，即 n 是随机变量。

考虑总体 $N(\mu, \sigma^2)$中均值 μ 的区间估计

1. 如果总体方差已知 $\sigma^2 = \sigma_0^2$

这时只要样本容量 n 足够大就可以找到同时满足置信度与长度要求的区间估计：

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

区间的长度为：$2u_{\alpha/2}\sigma_0/\sqrt{n} \leq l$

只要取样本容量 $n \geq \left(\frac{2u_{\alpha/2}\,\sigma_0}{l}\right)^2$ ，

就可以保证置信度与精度同时达到要求。

2. 如果总体方差 σ^2未知

1940年 Dantzig 证明了这种同时满足置信度 $\geq 1-\alpha$ 与长度 $\leq l$ 的区间估计不存在。

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

1945年，C.Stein 提出了一个两阶段抽样的序贯方法，能够同时满足置信度与长度的要求：

第一阶段先抽取若干样本估计 σ^2，根据这个估计值看需要多大的样本容量 n 才能满足长度的要求，不够的 n 将在第二阶段补齐。

给定一个自然数 n_0，记：$C = \dfrac{l^2}{4\,t^2_{\alpha/2}\,(n_0 - 1)}$，

$$\overline{X}_0 = \frac{1}{n_0}\sum_{k=1}^{n_0} X_k \qquad S^2 = \frac{1}{n_0 - 1}\sum_{k=1}^{n_0}(X_k - \overline{X}_0)^2$$

指定自然数 n_0，第一阶段抽样 n_0 次得到样本 X_1，…，X_{n_0}，定义：

$$n(t) = \max\left(n_0,\ \frac{t^2}{C} + 1\right),\ 0 < t < +\infty$$

得到 n_0 个样本后计算出样本标准差 S，如果函数$n(S) = n_0$ 即 $S^2 \le C\,n_0$ 则停止抽样 (长度已满足要求)；

另一种情况是 $n(S) > n_0$ 即 $S^2 > C n_0$，则进入第二阶段，再抽样 $n(S) - n_0$ 次，又得到样本 X_{n_0+1}，…，$X_{n(S)}$，定义随机变量：

$$\overline{X} = \frac{1}{n(S)} \sum_{k=1}^{n(S)} X_k \qquad Y = \frac{\sqrt{n(S)}(\overline{X} - \mu)}{S}$$

在理论上可以证明 $Y \sim t(n_0 - 1)$，所以如下区间具有置信度 $1 - \alpha$：

$$\left(\overline{X} - t_{\alpha/2}(n_0 - 1) \frac{S}{\sqrt{n(S)}}, \overline{X} + t_{\alpha/2}(n_0 - 1) \frac{S}{\sqrt{n(S)}} \right)$$

注意到函数 $n(t)$ 的定义，始终有 $n(S) \geq S^2/C$，因此这个区间的长度 $\leq l$ □

例2.3.7　假定总体 $X \sim N(\mu, \sigma^2)$，如下构造一个置信度 ≥ 0.95，长度 ≤ 0.6 的关于 μ 的区间估计。

第一阶段，假定取了 $n_0 = 21$ 个样本，

0.72	1.69	0.82	0.21	-0.09	1.57	-0.26
1.08	1.94	1.60	1.19	1.04	2.74	1.45
0.27	-0.84	0.82	0.77	1.93	0.36	0.13

这21个样本的样本均值与样本方差为：

$$\overline{X}_0 = 0.91, \quad S^2 = 0.73$$

相应的，

分位点 $t_{0.025}(20) = 2.086$，$C=0.6^2/(4\times 2.086^2)$，

$n(S)=\max(21, [0.73\times 4\times 2.086^2/0.6^2]+1) = 36$

因此需要再抽取 $n(S) - n_0 = 36 - 21 = 15$ 个样本，

1.23	2.11	1.07	1.02	1.84	-0.26	0.05	1.83
0.46	-0.06	0.69	2.32	1.05	0.44	0.30	

计算出全部 36 个样本的总的样本均值 0.92，以及相应的抽样误差$2.086\times 0.73^{1/2}/36^{1/2} = 0.297$，最后得到置信度 0.95，长度不超过 0.6 的区间：

(0.623，1.217)

□

练习2.3.8

假定成年人脉搏次数(每分钟)服从正态分布。

(1) 分别抽取 15 个男生样本、10 个女生样本，给出男、女生脉搏次数 0.95 的区间估计。

(2) 合并这 25 个样本计算 0.95 的区间估计。

(3) 给出 0.95 的上侧以及下侧区间。

(4) 讨论序贯区间，要求区间长度不大于 10。