参数估计

主讲人：刘宏志
liuhz@ss.pku.edu.cn

参数估计的动机

- 实际中，经常碰到研究总体的分布类型已知，但其依赖的（一个或几个）参数未知
  - 例：在单位时间间隔内某电话交换台接到呼叫的次数X是一个随机变量，由泊松流的性质推知其服从泊松分布。问题：单位时间间隔内接到k次呼叫的概率。
  - 例：某灯泡厂在稳定地条件下生产的灯泡的寿命是一个随机变量，由实际经验知道其服从$N(\mu,\sigma^2)$分布。问题：要评估该厂生产的灯泡的质量。
- 一些问题中，人们并不关心总体分布的形式，而只关心知道某些数字特征（如均值和方差）

如何估计这些已知分布的参数或未知分布的数字特征？

参数估计的基本思想

- 用所获得的样本值去估计参数取值

总体 $\xrightarrow{\text{抽取部分观察单位}}$ 样本

样本 → 统计量

参数 $\xleftarrow{\text{参数估计}}$ 统计量

参 数

如：总体均数 μ

总体标准差 σ

总体率 π

统计量

如：样本均数 $\bar{X}$

样本标准差 S

样本率 P

抽样调查

- 基本思想：
  - 一种科学的非全面的调查方法
  - 按照随机原则从调查对象总体中抽取部分单位进行调查，并根据部分单位的调查结果推断总体的数字特征
- 优点：
  - 抽样单元的随机选取排除了调查者的偏见（准确性）
  - 与完全枚举相比，小样本能减少成本（经济性）
  - 收集、整理数据、综合样本的速度快（时效性）
  - .....

抽样调查：理论依据

- 大数定律：随机现象出现的基本规律
  - 尽管观察过程中每次取得的结果不同（随机性），但大量重复观察结果的平均数却几乎接近某个确定的数值
  - 虽然样本误差在所难免，但只要样本容量足够大，计算出来的样本统计量就和总体参数非常接近

掷币次数	出现正面次数	出现正面的机会
4040	2048	0.5069
12000	6019	0.5016
24000	12012	0.5005

抽样：基本术语

- **总体**：抽样调查中所有调查对象的集合体
- **样本**：从总体中按一定方式抽取出的一部分元素的集合体
- **抽样**：从总体中按一定方式选择或抽取样本的过程
- **总体参数**：关于总体中某一变量的综合描述
- **样本统计量**：关于样本中某一变量的综合描述
- **抽样误差**：样本统计值和总体参数值之间的差异
- **置信度**：总体参数值落在样本统计值某一区间内的概率
  - 反映的是抽样的可靠性程度
- **置信区间**：给定置信度，样本统计值与总体参数值之间的误差范围
  - 反映的是抽样的准确性程度

北京大学

案例：医院出院人数

- 总体：由N=393个短期居留医院组成
- 令x_i表示1968年1月份第i个医院的出院人数
- 总体参数：
  - 总体均值：

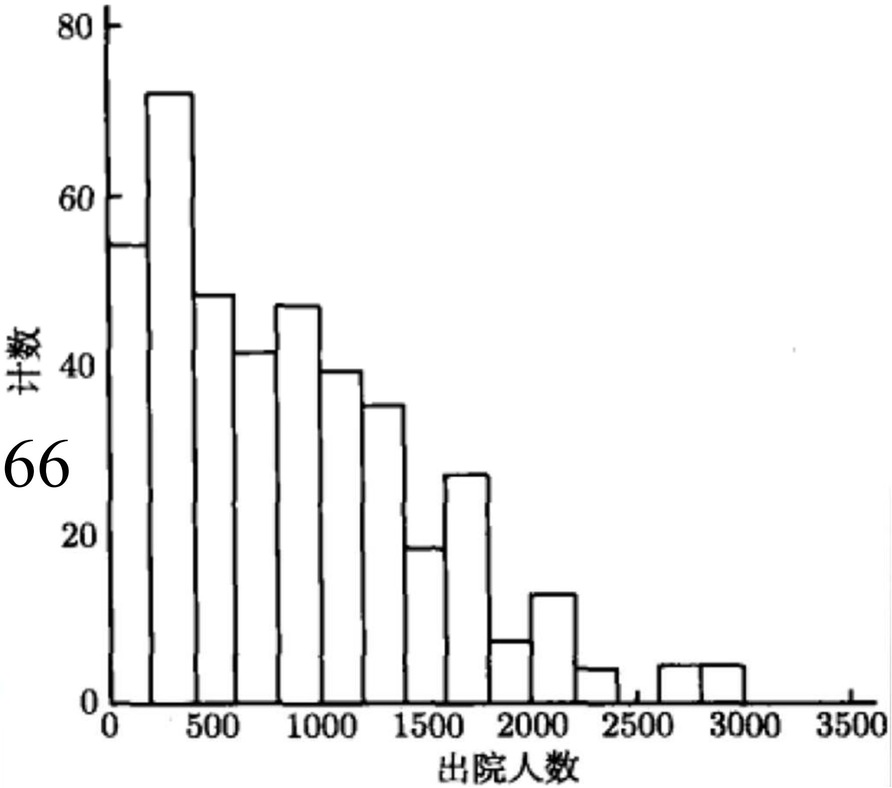
$$\mu = \frac{1}{N}\sum_{i=1}^{n} x_i = 814.6$$

  - 总体总数：

$$\tau = \sum_{i=1}^{n} x_i = N\mu = 320138$$

  - 总体方差：

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{n}(x_i - \mu)^2 = 347766$$



样本统计量

- 样本统计量是样本的函数，是随机变量，其分布称为抽样分布
  - 随机样本得到的数值或统计量都是随机的
  - 样本均值：$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$
  - 样本总数：$T=N\bar{X}$
  - 样本方差：$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$
- 例：以393个医院为总体，利用容量为n的样本进行参数估计

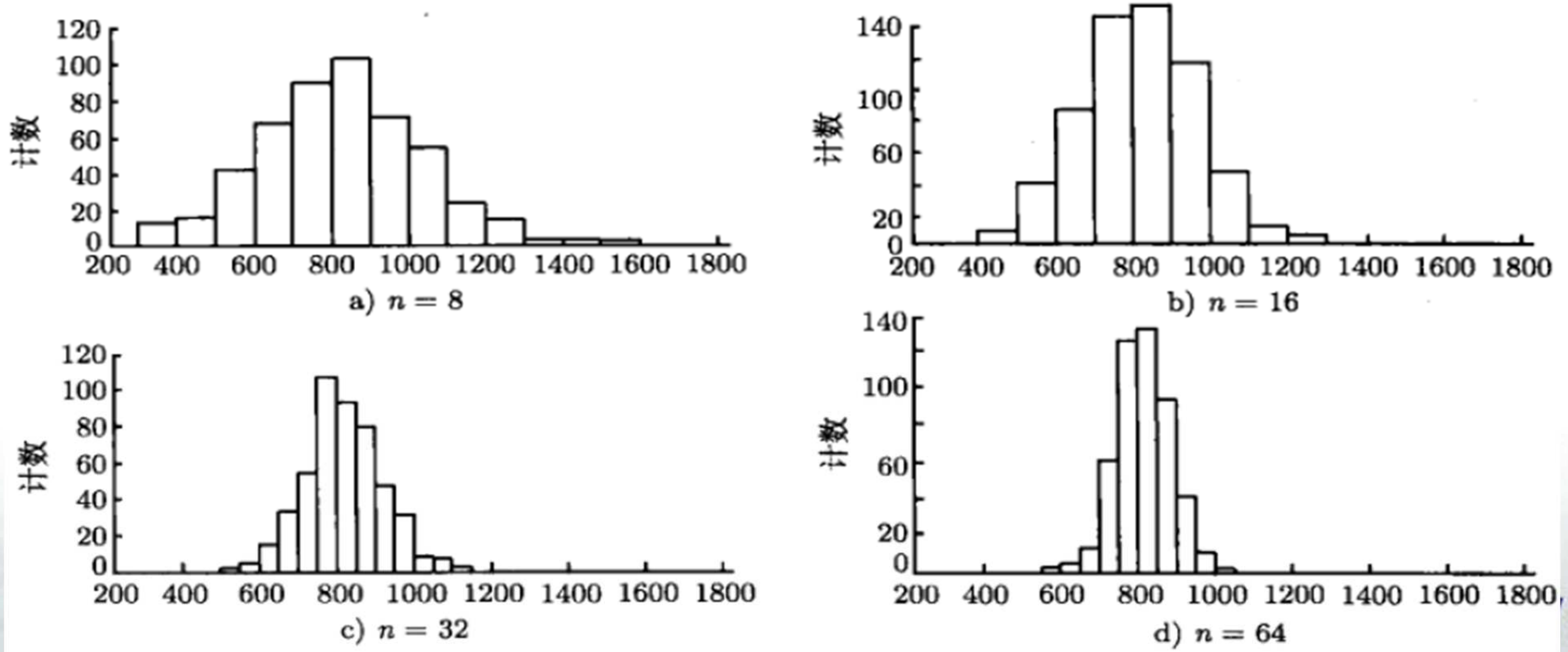


图 7.2　自 393 个医院总体中进行 500 次简单随机抽样时，不同样本容量下出院人数均值的直方图

参数估计方法分类

参数估计

- 点估计：用某一数值作为参数的近似值
- 区间估计：在要求的精度范围内指出参数所在的区间

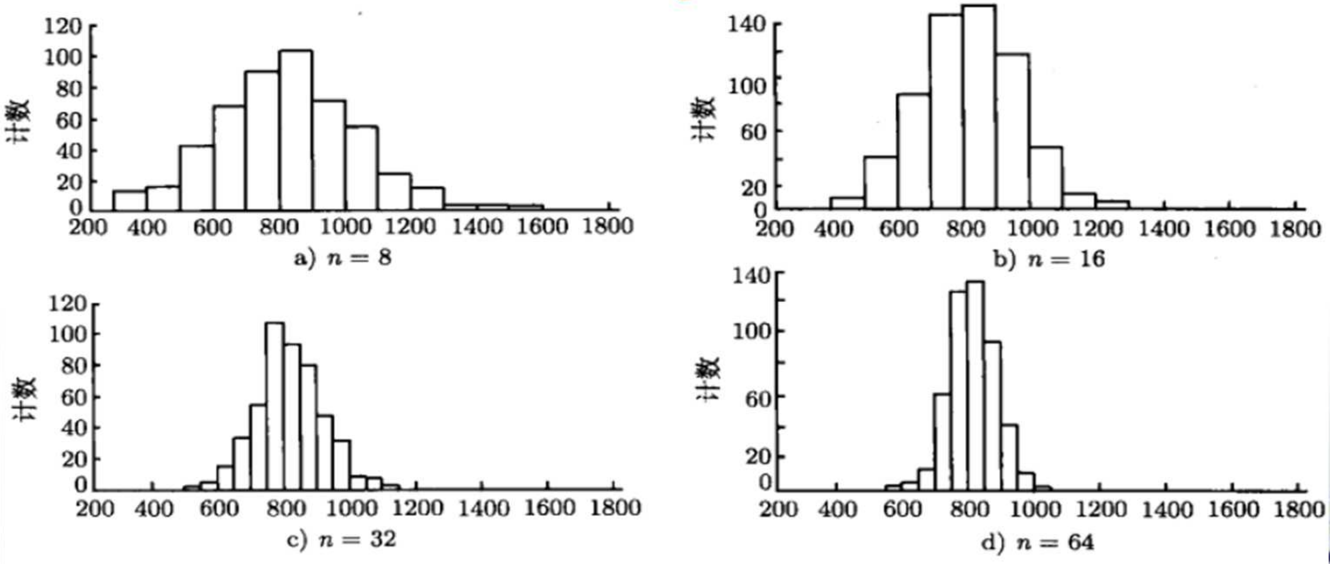


图 7.2 自 393 个医院总体中进行 500 次简单随机抽样时，不同样本容量下出院人数均值的直方图

点估计

- 设θ是总体X的未知参数
- 利用样本$X_1,X_2,\ldots,X_n$构造一个统计量$\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$来估计$\theta$
- $\hat{\theta}(X_1,X_2,\cdots,X_n)$称为$\theta$的点估计量，是一个随机变量
- 将样本观测值$(x_1,x_2,\ldots,x_n)$代入估计量$\hat{\theta}(X_1,X_2,\cdots,X_n)$，得到一个具体数值$\hat{\theta}(x_1,x_2,\cdots,x_n)$，称为$\theta$的点估计值
- 如果总体X有m个未知参数需要估计，则需要构造m个统计量分别作为对每个参数的估计
- 常用方法：矩估计法、最大似然法

矩估计法

- 基本思想：简单的“替换”
  - 用样本矩估计总体矩
- 原点矩
  - 定义：随机变量X的k次幂的数学期望(k为正整数)称为随机变量X的k阶原点矩：$\mu_k(X) = E(X^k)$
  - 一阶原点矩即为数学期望：$\mu_1(X) = E(X)$
- 中心矩
  - 定义：随机变量X的离差的k次幂的数学期望(k为正整数)称为随机变量X的k阶中心矩：$\gamma_k(X) = E\{[X - E(X)]^k\}$
  - 一阶中心矩恒等于零：$\gamma_1(X)$ =0
  - 二阶中心矩即为方差：$\gamma_2(X) = D(X)$
  - 三阶中心矩$E\{[X - E(X)]^3\}$主要衡量随机变量的分布是否有偏

北京大学

矩估计法

- 基本思想：简单的“替换”
  - 用样本矩估计总体矩
- 设 $(X_1, X_2, \ldots, X_n)$ 是来自总体 X 的一个样本, 根据大数定律, 对任意 $\varepsilon>0$, 有

$$\lim_{n\to\infty} P\{|\bar{X}-E(X)|\geq\varepsilon\}=0$$

- 对于任何 k, 只要 $E(X^k)$ 存在, 同样有

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n}X_i^k-E(X^k)|\geq\varepsilon\}=0, \qquad k=1,2,\ldots$$

矩估计法

设
$$\begin{cases} E(X)=\nu_1(\theta_1,\theta_2,\ldots,\theta_k)=\mu_1 \\ E(X^2)=\nu_2(\theta_1,\theta_2,\ldots,\theta_k)=\mu_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ E(X^k)=\nu_k(\theta_1,\theta_2,\ldots,\theta_k)=\mu_k \end{cases}$$

得到含有未知参数($\theta_1, \cdots, \theta_k$)的$k$个方程

解这k个联立方程组可得到($\theta_1, \cdots, \theta_k$)的一组解:

$$\begin{cases} \hat{\theta}_1=\hat{\theta}_1(X_1,X_2,\ldots,X_n) \\ \hat{\theta}_2=\hat{\theta}_2(X_1,X_2,\ldots,X_n) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \hat{\theta}_k=\hat{\theta}_k(X_1,X_2,\ldots,X_n) \end{cases}$$

用上面的解来估计参数 θ_i就是矩估计.

北京大学

例： 设某炸药厂一天中发生着火现象的次数X服从参数为λ的泊松分布，λ未知，有以下样本值；试估计参数λ。

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

解 $E(X) = \lambda \quad \mu_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$

令 $\bar{X} = \lambda,$

则 $\hat{\lambda} = \bar{x} = \frac{1}{250}(0 \times 75 + 1 \times 90 + \cdots + 6 \times 1) = 1.22$

所以 $\bar{X} = \lambda,$ 估计值$\hat{\lambda} = 1.22$。

最大似然估计

- 基本思想：

  根据样本值来选择参数，使该样本发生的概率最大

- **案例1：**某位同学与一位猎人一起外出打猎。一只野兔从前方窜过。只听一声枪响，野兔应声倒下。如果要你推测，是谁打中的？你会如何想?

分析：只发一枪便打中，猎人命中的概率大于这位同学命中的概率。

结论：看来这一枪是猎人射中的。

最大似然估计

- 基本思想：

  根据样本值来选择参数，使该样本发生的概率最大

- **案例2**：外形相同的两个箱子，甲箱有99个白球1个黑球，乙箱有1个白球99个黑球。随机地取一箱，并从中任取一球，结果发现是白球。问这球是从哪个箱中取出？

分析：若是从甲箱中取的，则取得白球的概率是99%；若是从乙箱中取的，则取得白球的概率为1%。

结论：这球应该是甲箱中取出的

似然函数

定义: 设总体 X 的分布类型已知,但含有未知参数 θ.

(1) 设离散型总体 X 的概率分布律为 $p(x;\theta)$，则样本 $(X_1,X_2,\cdots,X_n)$ 的联合分布律

$$p(x_1;\theta)p(x_2;\theta)\cdots p(x_n;\theta)=\prod_{i=1}^{n}p(x_i;\theta)$$

称为似然函数，并记之为 $L(\theta)=L(x_1,x_2,\cdots,x_n;\theta)=\prod_{i=1}^{n}p(x_i;\theta)$.

(2) 设连续型总体 X 的概率密度函数为 $f(x;\theta)$，则样本 $(X_1,X_2,\cdots,X_n)$ 的联合概率密度函数

$$f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta)=\prod_{i=1}^{n}f(x_i;\theta)$$

仍称为似然函数，并记之为 $L(\theta)=L(x_1,x_2,\cdots,x_n;\theta)=\prod_{i=1}^{n}f(x_i;\theta)$.

极大似然估计值/量

定义：设总体的分布类型已知,但含有未知参数 θ.

(1) 设 $(x_1,x_2,\cdots,x_n)$ 为总体 X 的一个样本观察值，若似然函数 $L(\theta)$ 在 $\hat{\theta}=\hat{\theta}(x_1,x_2,\cdots,x_n)$ 处取到最大值，则称 $\hat{\theta}(x_1,x_2,\cdots,x_n)$ 为 θ 的极大似然估计值.

(2) 设 $(X_1,X_2,\cdots,X_n)$ 为总体 X 的一个样本，若 $\hat{\theta}(x_1,x_2,\cdots,x_n)$ 为 θ 的极大似然估计值，则称 $\hat{\theta}(X_1,X_2,\cdots,X_n)$ 为参数 θ 的极大似然估计量.

最大似然估计

一般步骤：

（1）求似然函数 $L(\theta)$；

（2）求出 $\ln L(\theta)$ 及方程 $\frac{d}{d\theta}\ln L(\theta)=0$；

（3）解上述方程得到极大似然估计值

$$\hat{\theta}=\hat{\theta}(x_1,x_2,\cdots,x_n).$$

（4）解上述方程得到极大似然估计量

$$\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n).$$

案例：池塘鱼数估计

- 池中有许多鱼，捉住500条，做上标记再放入池中，待充分混合后，再捉1000条鱼，其中100条鱼带记号。试估计池中有多少条鱼？
- **分析：**设池中鱼的总条数（待估计量）为N，其中r条鱼有标记，随机捉s条鱼发现有x条带标记；用X记捉住s条鱼中带标记的鱼数

$$P(X=x)=\binom{N-r}{s-x}\binom{r}{x}/\binom{N}{s}$$

- 似然函数$L(N)=P(X=x)$

$$\frac{L(N)}{L(N-1)}=\frac{(N-s)(N-r)}{(N+x-r-s)N}=\frac{N^2-(r+s)N+rs}{N^2-(r+s)N+xN}$$

- 当$rs>xN$时，$L(N)>L(N-1)$;当$rs<xN$时，$L(N)<L(N-1)$
- 故N的极大似然估计量为$\hat{N}=\left[\frac{rs}{X}\right]$，代入数值的$\hat{N}$ =5000

北京大学

估计量的评选标准

- 动机：
  - 对同一参数，用不同方法可得到不同的估计量
  - 同一参数的多个估计量，哪个更好？
- 评价标准：估计量的统计性质
  - 无偏性
  - 有效性
  - 一致性

无偏性

- 动机：
  - 估计量是随机变量，不同样本值会得到不同估计值
  - 希望估计值在真值附近波动，且其期望等于未知参数的真实值
- 定义：

  设$\hat{\theta}(X_1, X_2, \ldots, X_n)$为$\theta \in \Theta$的估计量.

  若 $E[\hat{\theta}(X_1, X_2, \ldots, X_n)] = \theta, \quad \forall \theta \in \Theta$

  则称$\hat{\theta}(X_1, X_2, \ldots, X_n)$为$\theta$的无偏估计. 否则称为有偏的
- 含义：$E(\hat{\theta}) = \theta$
  - $\hat{\theta}$的平均值恰好等于θ的真值，即要求没有系统误差
  - 用一台秤去称物体，误差源有两个：一是秤本身制作结构上的问题，属于系统误差；另一种是操作或其它随机因素的干扰，属于随机误差

例: 设总体X的k阶矩$E(X^k)$存在,证明样本的k阶矩是$E(X^k)$的无偏估计.

证明 因为

$$E\left(\frac{1}{n}\sum_{i=1}^{n} X_i^k\right) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i^k\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i^k)$$

$$= \frac{1}{n}\sum_{i=1}^{n} E(X^k) = E(X^k)$$

所以,证明样本的k阶矩是$E(X^k)$的无偏估计.

例: 设总体的方差$D(X)$存在,试证样本二阶中心矩γ_2是总体方差$D(X)$的有偏估计.

证明

$$E(\gamma_2)=E[\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2]=E[\frac{1}{n}\sum_{i=1}^{n}X_i^2-(\bar{X})^2]$$

$$=E(\frac{1}{n}\sum_{i=1}^{n}X_i^2)-E[(\bar{X})^2]=E(X^2)-E[(\bar{X})^2]$$

$$=D(X)+[E(X)]^2-D(\bar{X})-[E(\bar{X})]^2$$

$$=D(X)-\frac{1}{n}D(X)=\frac{n-1}{n}D(X)$$

$$\begin{aligned} D(\bar{X}) &= D(\tfrac{1}{n}\textstyle\sum_{i=1}^{n} X_i) \\ &= \tfrac{1}{n^2}D(\textstyle\sum_{i=1}^{n} X_i) \\ &= \tfrac{1}{n^2}(\textstyle\sum_{i=1}^{n} D(X_i)) \\ &= \tfrac{\sigma^2}{n} \end{aligned}$$

所以, γ_2是总体方差$D(X)$的有偏估计.

$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$是$D(X)$的无偏估计.

北京大学

有效性

- 动机：
  - 无偏性只保证了估计量取值在参数真值周围波动
  - 但未考虑波动幅度的大小
  - 一个参数的无偏估计量不是唯一的
  - 希望波动的幅度越小越好
  - 方差是随机变量取值与其期望的偏离程度的度量
- 定义：

设$\hat{\theta}_1=\hat{\theta}_1(X_1,X_2,\ldots,X_n)$与$\hat{\theta}_2=\theta_2(X_1,X_2,\ldots,X_n)$

都是θ的无偏估计量，若对$\forall\theta\in\Theta$有$D(\hat{\theta}_1)\le D(\hat{\theta}_2)$，且至少

有一个$\theta\in\Theta$使不等式成立，则称$\hat{\theta}_1$ 比$\hat{\theta}_2$有较高的效率，简

称$\hat{\theta}_1$比$\hat{\theta}_2$<u>有效</u>.

例: 设(X_1,X_2, X_3)是来自总体X的一个样本,证明下面的三个估计量都是总体均值$E(X)$的无偏估计量

$$\hat{\mu}_1 = \overline{X};\quad \hat{\mu}_2 = \frac{1}{2}X_1 + \frac{1}{3}X_2 + \frac{1}{6}X_3;\quad \hat{\mu}_3 = X_1$$

且$\hat{\mu}_1$较$\hat{\mu}_2,\hat{\mu}_3$都有效.

证明 显然有 $E(\hat{\mu}_1) = E(\hat{\mu}_2) = E(\hat{\mu}_3)$

且 $D(\hat{\mu}_1) = D(\overline{X}) = D(X)/3$

$$D(\hat{\mu}_2) = D(X_1/2 + X_2/3 + X_3/6) = 14D(X)/36$$

$$D(\hat{\mu}_3) = D(X_1) = D(X)$$

故有$D(\hat{\mu}_1) < D(\hat{\mu}_2) < D(\hat{\mu}_3)$,所以$\hat{\mu}_1$较$\hat{\mu}_2,\hat{\mu}_3$有效.

一致性

- 动机：
  - 无偏性和有效性都是在样本容量固定的前提下提出的
  - 希望随着样本容量增大,估计值稳定于待估参数的真值
- 定义：

设总体X有概率函数$p(x;\theta),\theta\in\Theta$为待估参数,

$\hat{\theta}_n(X_1,X_2,\ldots,X_n)$为$\theta$的估计量.若对于任意$\varepsilon>0$,总有

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

则称$\hat{\theta}_n(X_1,X_2,\ldots,X_n)$为参数$\theta$的<u>一致估计</u>.

关于一致性的常用结论

- 样本的k阶矩是总体的k阶矩的一致性估计量
  - 利用大数定律可证明
- 设$\hat{\theta}$是θ的无偏估计量，且$\lim\limits_{n\to\infty} D(\hat{\theta})=0$，则$\hat{\theta}$是$\theta$的一致估计量
  - 利用切比雪夫不等式可证明

矩估计和极大似然估计得到的估计量一般为一致估计量

区间估计

- 动机：点估计不能反映估计的误差和精确程度
  - 区间估计利用样本统计量和抽样分布估计总体参数的可能区间
- 抽样误差：
  - 一个无偏估计与其对应的总体参数之差的绝对值。
  - 例：均值的抽样误差 $e = |\bar{x} - \mu|$（实际未知）
  - 区间估计的关键是对抽样误差e进行求解
  - 若e已知，则均值的区间可表示为：$[\bar{x} - e, \bar{x} + e]$
- 正态分布的标准方差σ已知时，均值的区间估计：$\bar{x} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$
  - $(1-\alpha)$为置信系数
  - $Z_{\alpha/2}$为在标准正态分布的右侧尾部所提供的面积为 $\alpha/2$ 的Z值
- 标准方差σ未知时，均值的区间估计为：$\bar{x} \pm Z_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$

北京大学

区间估计：案例

斯泰特怀特保险公司每年都需对人寿保险单进行审查，现公司抽取36个寿保人作为一个简单随即样本，得到关于、投保人年龄、保费数量、保单的现金值、残废补偿选择等项目的资料。为了便于研究，某位经理要求了解寿险投保人总体平均年龄的90%的区间估计。

投保人	年龄	投保人	年龄	投保人	年龄	投保人	年龄
1	32	10	47	19	27	28	34
2	50	11	31	20	43	29	39
3	40	12	36	21	54	30	34
4	24	13	39	22	36	31	35
5	33	14	46	23	34	32	42
6	44	15	45	24	48	33	53
7	45	16	39	25	23	34	28
8	48	17	38	26	36	35	49
9	44	18	45	27	42	36	39

区间估计：案例

上表是一个由36个投保人组成的简单随机样本的年龄数据。现求总体的平均年龄的区间估计。

分析：区间估计包括两个部分——点估计和误差边际，只需分别求出即可到的总体的区间估计。

解：已知 $n=36$（大样本）， $1-\alpha=90\%$， $Z_{\alpha/2}=1.645$

（1）样本的平均年龄 $\bar{x}=\dfrac{\sum x}{n}=\dfrac{32+50+40+\cdots+36}{36}=39.5$

（2）误差边际 $E=Z_{\alpha/2}\cdot\dfrac{\sigma}{\sqrt{n}}\rightarrow$ 总体标准差σ（未知）

$\rightarrow$ 样本标准差s

区间估计：案例

样本标准差 $s=\sqrt{\dfrac{\sum(x-\bar{x})^2}{n-1}}=7.77$

误差边际 $E=Z_{\alpha/2}\cdot\dfrac{\sigma}{\sqrt{n}}=Z_{\alpha/2}\cdot\dfrac{s}{\sqrt{n}}=1.645*\dfrac{7.77}{\sqrt{36}}=2.13$

（3）90%的置信区间为39.5 ±2.13　即（37.37，41.63）岁。

注意

置信区间的长度（准确度）在置信度一定的情况下，与样本容量的大小呈反方向变动，若要提高估计准确度，可扩大样本容量