第六章　点估计(Point Estimation)

第一节　矩法估计

第二节　极大似然估计

第三节　克拉默-拉奥（Cramer-Rao）不等式

第四节　充分统计量

第五节　拉奥-勃拉克维（Rao-Blackwell）定理和一致最小方差无偏估计

§6.1 矩法估计

Method of Moments Estimation

一、点估计问题的提法

设总体 X 的分布函数形式已知, 但它的一个或多个参数为未知, 借助于总体 X 的一个样本来估计总体未知参数的值的问题称为点估计问题.

例6.1.1 在某炸药制造厂, 一天中发生着火现象的次数 X 是一个随机变量, 假设它服从以 $\lambda > 0$ 为参数的泊松分布, 参数 λ 为未知, 设有以下的样本值, 试估计参数 λ.

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

解 因 $X \sim P(\lambda)$, 所以 $\lambda = E(X)$.

用样本均值来估计总体的均值 $E(X)$.

$$\bar{x} = \frac{\sum_{k=0}^{6} k n_k}{\sum_{k=0}^{6} n_k} = \frac{1}{250}(0\times 75 + 1\times 90 + 2\times 54 + 3\times 22 + 4\times 6 + 5\times 2 + 6\times 1) = 1.22.$$

故 $E(X) = \lambda$ 的估计为 1.22 .

点估计问题的一般提法

设总体 X 的分布函数 $F(x;\theta)$ 的形式为已知, θ 是待估参数. $X_1, X_2, \cdots, X_n$ 是 X 的一个样本, $x_1, x_2, \cdots, x_n$ 为相应的一个样本值.

点估计问题就是要构造一个适当的统计量 $\hat{\theta}(X_1, X_2, \cdots, X_n)$, 用它的观察值 $\hat{\theta}(x_1, x_2, \cdots, x_n)$ 来估计未知参数 θ.

$\hat{\theta}(X_1, X_2, \cdots, X_n)$ 称为 θ 的估计量.
$\hat{\theta}(x_1, x_2, \cdots, x_n)$ 称为 θ 的估计值.
} 通称点估计, 简记为 $\hat{\theta}$.

二、估计量的求法

常用构造估计量的方法: (两种)

矩估计法和极大似然估计法.

1. 矩估计法

基本概念

设 $X_1, X_2, \cdots, X_n$ 是来自总体X的样本,

若 $E(X^k)$, $k = 1,2,\cdots$存在,

称它为总体X 的 k 阶原点矩,简称 k 阶矩.

样本 k 阶(原点)矩 $$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k, \ k = 1, 2, \cdots;$$

其观察值 $$a_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k, \ k = 1, 2, \cdots.$$

结论：若总体 X 的 k 阶矩 $E(X^k) \xlongequal{\text{记成}} \mu_k$ 存在,

则当 $n \to \infty$ 时, $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{P} \mu_k, k = 1, 2, \cdots.$

证明：因为 $X_1, X_2, \cdots, X_n$ 独立且与 X 同分布,

所以 $X_1^k, X_2^k, \cdots, X_n^k$ 独立且与 X^k 同分布,

故有 $E(X_1^k) = E(X_2^k) = \cdots = E(X_n^k) = \mu_k.$

再根据第四章辛钦大数定律知

$$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k \xrightarrow{P} \mu_k\,,\qquad k=1,2,\cdots;$$

由第四章关于依概率收敛的序列的性质知

$$\boxed{g(A_1,A_2,\cdots,A_k)\xrightarrow{P} g(\mu_1,\mu_2,\cdots,\mu_k),}$$

其中$g(x_1,x_2,\cdots,\ x_n)$是有理函数，且$g(\mu_1,\mu_2,\cdots,\ \mu_n)\neq\pm\infty.$

这一结论是接下来所要介绍的矩估计法的理论根据.

设 X 为连续型随机变量,其概率密度为 $f(x;\theta_1,\theta_2,\cdots,\theta_k)$, 或 X 为离散型随机变量, 其分布律为 $P\{X=x\}=p(x;\theta_1,\theta_2,\cdots,\theta_k)$, 其中 $\theta_1,\theta_2,\cdots,\theta_k$ 为待估参数,

若 $X_1,X_2,\cdots,X_n$ 为来自总体X 的样本,

假设总体 X 的r阶原点矩存在,

且均为 $\theta_1,\theta_2,\cdots,\theta_k$ 的函数, 即

$$E(X^r)=\mu_r=\mu_r(\theta_1,\theta_2,\cdots,\theta_k)\quad r=1,2,\cdots,k$$

矩估计法思想:

用样本的r阶原点矩 $A_r=\frac{1}{n}\sum_{i=1}^{n}X_i^r$ 来估计总体的r阶原点矩 $\mu_r=E(X^r),r=1,2,\cdots,k.$

具体做法: 考虑1~k阶矩，得方程组:

令
$$\begin{cases} E(X)=\frac{1}{n}\sum_{i=1}^{n}X_i \\ E(X^2)=\frac{1}{n}\sum_{i=1}^{n}X_i^2 \\ \cdots \\ E(X^k)=\frac{1}{n}\sum_{i=1}^{n}X_i^k \end{cases}$$

根据总体的分布计算总体的k阶矩

$$\begin{cases} \mu_1(\theta_1,\theta_2,\cdots,\theta_k)=\frac{1}{n}\sum_{i=1}^{n}X_i \\ \mu_2(\theta_1,\theta_2,\cdots,\theta_k)=\frac{1}{n}\sum_{i=1}^{n}X_i^2 \\ \cdots \\ \mu_k(\theta_1,\theta_2,\cdots,\theta_k)=\frac{1}{n}\sum_{i=1}^{n}X_i^k \end{cases}$$

这是一个包含k个未知量 $\theta_1, \theta_2, \cdots, \theta_k$ 的方程组,

于是我们可以解方程,

用方程组的解 $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k$ 分别作为 $\theta_1, \theta_2, \cdots, \theta_k$ 的估计量,这个估计量称为矩估计量.

矩估计量的观察值称为矩估计值.

例6.1.2 设总体X在$[0,\theta]$上服从均匀分布,其中θ $(\theta>0)$未知,$(X_1,X_2,\cdots,X_n)$是来自总体X的样本,求θ的矩估计量.

解 因$E(X)=\dfrac{\theta}{2}$,

根据矩估计法,令$E(X)=\dfrac{1}{n}\sum_{i=1}^{n}X_i=\bar{X}$, 即 $\dfrac{\theta}{2}=\bar{X}$,

解得 $\hat{\theta}=2\overline{X}$ 为所求θ的矩估计量.

例6.1.3

$$r.v.X \sim f(x)=\begin{cases}(\theta+1)x^{\theta} & 0<x<1 \\ 0 & \text{其它}\end{cases}$$，其中$\theta>-1$未知，

求θ的矩估计量。

解

$$E(X)=\int_{-\infty}^{+\infty} xf(x)dx=\int_0^1 x(\theta+1)x^{\theta}dx$$

$$=\int_0^1(\theta+1)x^{\theta+1}dx=\frac{\theta+1}{\theta+2}$$

令$E(X)=A_1$，即$\frac{\theta+1}{\theta+2}=\bar{X}$

解之得θ的矩估计量为 $\hat{\theta}=\frac{2\bar{X}-1}{1-\bar{X}}$

例6. 1. 4 设总体 X 在$[a,b]$上服从均匀分布,其中a, b未知, $(X_1, X_2,\cdots, X_n)$是来自总体X的样本,求a, b的矩估计量.

解 因为 $\mu_1 = E(X)=\dfrac{a+b}{2}$,

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

令

$$\begin{cases} \dfrac{a+b}{2}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i = \bar{X} = A_1 \\ \dfrac{(a-b)^2}{12}+\dfrac{(a+b)^2}{4}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^2 = A_2 \end{cases}$$

即 $\begin{cases} a+b=2A_1, \\ b-a=\sqrt{12(A_2-A_1{}^2)}. \end{cases}$

解方程组得到*a*, *b*的矩估计量分别为

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

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

例6.1.5 设总体 X 的均值 μ 和方差 σ^2 都存在,且有 $\sigma^2>0$, 但 μ 和 σ^2 均为未知, 又设 $X_1,X_2,\cdots,X_n$ 是一个样本, 求 μ 和 σ^2 的矩估计量.

解

$$E(X)=\mu,$$

$$E(X^2)=D(X)+[E(X)]^2=\sigma^2+\mu^2,$$

令 $\begin{cases}\mu=A_1,\\ \sigma^2+\mu^2=A_2.\end{cases}$

解方程组得到矩估计量分别为 $\hat{\mu}=A_1=\overline{X}$,

$$\hat{\sigma}^2=A_2-A_1^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2.$$

上例表明:

不管总体服从什么类型的分布,

用样本均值$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$作为总体均值的矩估计,

用样本二阶中心矩 $B_2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$作为总体方差的矩估计.

§6.2 极大似然估计

极大似然估计（ **Maximum Likelihood Estimation, MLE**）**,**也称为最大概似估计或最大似然估计，是求估计的另一种方法，**1821**年首先由德国数学家高斯（**C. F. Gauss**）提出，但是这个方法通常被归功于英国的统计学家罗纳德·费希尔（**R. A. Fisher**）。

一、极大似然估计法

极大似然思想：

引例 在一盒中放有红球和白球，且已知两种颜色的球的比例为1:9, 但不知何种球多, 一人从盒中有放回地随机抽取3个球, 发现都为红球, 试推断盒中红球多, 还是白球多?

分析：引进随机变量$X = \begin{cases} 1, & \text{取得的球为红球} \\ 0, & \text{否则} \end{cases}$

则X就看作一个总体,显然它服从$0-1$分布,

它有一个未知参数p, p只能取0.1或0.9

现在我们进行了抽样，得到一个容量为3的子样：1,1,1. 那么这个样本值出现的概率是

$$P(X_1=1, X_2=1, X_3=1)=p^3$$

当p=0.1时其值为0.001，而当p=0.9时其值为0.729.

显然此时我们认为p=0.9更合理.

极大似然原理：一个E如有若干个可能结果A, B, C, …，在一次试验中，结果A出现，则一般认为试验条件对A出现有利，即A出现的概率很大.

似然函数的定义

(1) 设总体 X 是离散型随机变量,

设分布律 $P\{X = x\} = p(x;\theta)$, θ 为待估参数, $\theta \in \Theta$,

(其中 Θ 是 θ 可能的取值范围)

$X_1, X_2, \cdots, X_n$是来自总体 X 的样本,

又设 $x_1, x_2, \cdots, x_n$ 为样本 $X_1, X_2, \cdots, X_n$ 的一个样本值.

则样本 $X_1, X_2, \cdots, X_n$ 取到观察值 $x_1, x_2, \cdots, x_n$ 的概率,

即事件 $X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n$ 发生的概率为

$$L(\theta) = L(x_1, x_2, \cdots, x_n; \theta)$$

$$= P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n) = \prod_{i=1}^{n} p(x_i; \theta), \quad \theta \in \Theta,$$

$L(\theta)$称为样本似然函数.

注意: 在似然函数$L(x_1, x_2, \cdots, x_n; \theta)$中, $x_1, x_2, \cdots, x_n$是已知样本观测值, 视作已知常数, 未知变量只有θ.

(2) 设总体X是连续型随机变量，

设概率密度为 $f(x;\theta)$, θ 为待估参数， $\theta \in \Theta$,

(其中 Θ 是 θ 可能的取值范围)

$X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本,

则 $X_1, X_2, \cdots, X_n$ 的联合密度为 $\prod_{i=1}^{n} f(x_i;\theta)$.

又设 $x_1, x_2, \cdots, x_n$ 为相应于样本 $X_1, X_2, \cdots, X_n$ 的一个样本值.

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

得到样本值 $x_1, x_2, \cdots, x_n$ 时,选取使似然函数 $L(\theta)$ 取得最大值的 $\hat{\theta}$ 作为未知参数 θ 的估计值,

即 $L(x_1, x_2, \cdots, x_n; \hat{\theta}) = \sup\limits_{\theta \in \Theta} L(x_1, x_2, \cdots, x_n; \theta).$

(其中 Θ 是 θ 可能的取值范围)

这样得到的 $\hat{\theta}$ 与样本值 $x_1, x_2, \cdots, x_n$ 有关,记为

$\hat{\theta}(x_1, x_2, \cdots, x_n)$, 参数 θ 的最大似然估计值,

(Maximum Likelihood Estimate)

$\hat{\theta}(X_1, X_2, \cdots, X_n)$ 参数 θ 的最大似然估计量.

(Maximum Likelihood Estimator)

求极大似然估计量的步骤:

(一) 写出似然函数

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

(二) 取对数

$$\ln L(\theta)=\sum_{i=1}^{n}\ln f(x_i;\theta);$$

(三) 对 θ 求导 $\dfrac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\theta}$, 并令 $\boxed{\dfrac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\theta}=0}$, 对数似然方程

解方程即得未知参数θ的最大似然估计值$\hat{\theta}$.

最大似然估计法也适用于分布中含有多个未知参数的情况. 此时只需令

$$\boxed{\frac{\partial}{\partial \theta_i} \ln L = 0, \qquad i = 1,2,\cdots,k.}$$ 对数似然方程组

解出由 k 个方程组成的方程组,即可得各未知参数 $\theta_i\ (i = 1,2,\cdots,k)$ 的最大似然估计值 $\hat{\theta}_i$.

例6.2.2 设 $X \sim B(1, p)$, $X_1, X_2, \cdots, X_n$ 是来自 X 的一个样本,求 $p\,(0 < p < 1)$ 的最大似然估计量.

解 设 $x_1, x_2, \cdots, x_n$ 为相应于样本 $X_1, X_2, \cdots, X_n$ 的一个样本值,

X 的分布律为 $P\{X = x\} = p^x(1-p)^{1-x}$, $x = 0,1$,

似然函数 $L(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$

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

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

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

由 $\frac{d^2}{\mathrm{d}p^2}\ln L(p)<0$ 知 p 的最大似然估计值为

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

p 的最大似然估计量为 $\hat{p}=\frac{1}{n}\sum_{i=1}^{n}X_i=\bar{X}.$

这一估计量与矩估计量是相同的.

例6.1.3(续)

$$r.v.X \sim f(x)=\begin{cases}(\theta+1)x^{\theta} & 0<x<1\\ 0 & \text{其它}\end{cases}，\text{其中}\theta>-1\text{未知，}$$

求 θ 的极大似然估计。

解 似然函数 $L(\theta)=\prod_{i=1}^{n}(\theta+1)x_i^{\theta}=(\theta+1)^n\left(\prod_{i=1}^{n}x_i\right)^{\theta}$

$$\ln L(\theta)=n\ln(1+\theta)+\theta\sum_{i=1}^{n}\ln x_i$$

令 $\dfrac{\mathrm{d}}{\mathrm{d}\theta}\ln L(\theta)=\dfrac{n}{1+\theta}+\sum_{i=1}^{n}\ln x_i=0,$

解之得 θ 的极大似然估计为 $\hat{\theta}=-\dfrac{n}{\sum_{i=1}^{n}\ln x_i}-1$

练习 设 X 服从参数为 $\lambda\ (\lambda > 0)$ 的泊松分布，$X_1, X_2, \cdots, X_n$ 是来自 X 的一个样本，求 λ 的最大似然估计量.

解 因为 X 的分布律为

$$P\{X = x\} = \frac{\lambda^x}{x!} e^{-\lambda} \quad (x = 0, 1, 2, \cdots, n, \cdots)$$

所以 λ 的似然函数为

$$L(\lambda) = \prod_{i=1}^{n} \left(\frac{\lambda^{x_i}}{x_i!} e^{-\lambda} \right) = e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} (x_i!)},$$

$$\ln L(\lambda) = -n\lambda + \left(\sum_{i=1}^{n} x_i\right)\ln\lambda - \ln\prod_{i=1}^{n}(x_i!),$$

令 $\dfrac{\mathrm{d}}{\mathrm{d}\lambda}\ln L(\lambda) = -n + \dfrac{\sum_{i=1}^{n}(x_i)}{\lambda} = 0,$

解得 λ 的最大似然估计值 $\hat{\lambda} = \dfrac{1}{n}\sum_{i=1}^{n} x_i = \bar{x},$

λ 的最大似然估计量为 $\hat{\lambda} = \dfrac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}.$

这一估计量与矩估计量是相同的.

例6.2.3 设总体 $X \sim N(\mu,\sigma^2)$, μ,σ^2为未知参数, $x_1, x_2, \cdots, x_n$ 是来自X的一个样本值, 求μ和σ^2的最大似然估计量.

解 X的概率密度为 $f(x;\mu,\sigma^2)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$,

X 的似然函数为

$$L(\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

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

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

令
$$\begin{cases} \dfrac{\partial}{\partial\mu}\ln L(\mu,\sigma^2) = 0, \\ \dfrac{\partial}{\partial\sigma^2}\ln L(\mu,\sigma^2) = 0, \end{cases}$$

$$\begin{cases} \dfrac{1}{\sigma^2}\left[\displaystyle\sum_{i=1}^{n} x_i - n\mu\right] = 0, \\ -\dfrac{n}{2\sigma^2} + \dfrac{1}{2(\sigma^2)^2}\displaystyle\sum_{i=1}^{n}(x_i-\mu)^2 = 0, \end{cases}$$

由 $\frac{1}{\sigma^2}\left[\sum_{i=1}^{n} x_i - n\mu\right] = 0$ 解得 $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$,

由 $-\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i - \mu)^2 = 0$ 解得

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

故 μ 和 σ^2 的最大似然估计量分别为

$\hat{\mu} = \overline{X}$, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$. 它们与相应的矩估计量相同.

$$L(x_1,x_2,\cdots,x_n;\hat{\theta})=\sup_{\theta\in\Theta} L(x_1,x_2,\cdots,x_n;\theta).$$

例6.2.4 设总体 X 在 $[a,b]$ 上服从均匀分布,其中 a, b 未知, $x_1,x_2,\cdots,x_n$ 是来自总体 X 的一个样本值,求 a,b 的最大似然估计量.

解 X 的概率密度为

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

因为$a \le x_1, x_2, \cdots, x_n \le b$等价于$a \le x_{(1)}, x_{(n)} \le b$,

作为a, b的函数的似然函数为

$$L(a,b)=\begin{cases}\dfrac{1}{(b-a)^n}, & a \le x_{(1)},\ x_{(n)} \le b, \\ 0, & \text{其他}\end{cases}$$

于是对于满足条件 $a \le x_{(1)},\ b \ge x_{(n)}$的任意$a,b$有

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

即似然函数 $L(a,b)$ 在 $a = x_{(1)}$, $b = x_{(n)}$ 时取到最大值 $(x_{(n)} - x_{(1)})^{-n}$,

a, b 的最大似然估计值

$$\hat{a} = x_{(1)} = \min_{1\le i\le n} x_i, \qquad \hat{b} = x_{(n)} = \max_{1\le i\le n} x_i,$$

a, b 的最大似然估计量

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

设总体X具有分布律：

X	1	2	3
P	θ^2	$2\theta(1-\theta)$	$(1-\theta)^2$

其中$0<\theta<1$ 未知。已知取得了一组样本观测值$x_1=1, x_2=2, x_3=1$, 试求θ 的矩估计值和极大似然估计值。

极大似然估计的不变性

设 θ 的函数 $u=u(\theta), \theta\in\Theta$ 具有单值反函数 $\theta=\theta(u), u\in U$ 又设 $\hat{\theta}$ 是 X 的概率密度函数 $f(x;\theta)$ (f 形式已知) 中的参数 θ 的最大似然估计, 则 $\hat{u}=u(\hat{\theta})$ 是 $u(\theta)$ 的最大似然估计 .

证明 因为$\hat{\theta}$是θ 的最大似然估计值，所以

$$L(x_1,x_2,\cdots,x_n;\hat{\theta})=\sup_{\theta\in\Theta}L(x_1,x_2,\cdots,x_n;\theta),$$

其中$x_1,x_2,\cdots,x_n$是来自总体 X的一个样本值，

由于

$$\hat{u} = u(\hat{\theta}), \qquad \hat{\theta} = \theta(\hat{u}),$$

故

$$L(x_1, x_2, \cdots, x_n; \theta(\hat{u})) = \sup_{u \in U} L(x_1, x_2, \cdots, x_n; \theta(u)),$$

于是 $\hat{u} = u(\hat{\theta})$是$u(\theta)$的最大似然估计.

此性质可以推广到总体分布中含有多个未知参数的情况.

在例6.2.3中, σ^2 的最大似然估计值为

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

函数 $u = u(\sigma^2) = \sqrt{\sigma^2}$ 有单值反函数

$$\sigma^2 = u^2 \quad (u \geq 0),$$

故标准差 σ 的最大似然估计量为

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

例6.2.5

设母体$\xi \sim \pi(\lambda)$, $\lambda > 0$未知. $\xi_1, \xi_2, \cdots, \xi_n$是来自$\xi$的子样, 试求$\theta_1 = e^{\lambda}$和$\theta_2 = 1/\lambda$的极大似然估计量.

注意到$y=e^x$在$x>0$上单调递增，因而有唯一的反函数．从而$\theta_1=e^\lambda$的极大似然估计量为

$$\hat{\theta}_1=e^{\hat{\lambda}}=e^{\bar{\xi}}.$$

注意到$y=1/x$在$x>0$上单调递减，因而有唯一的反函数．$\theta_2=1/\lambda$的极大似然估计量为

$$\hat{\theta}_2=\frac{1}{\hat{\lambda}}=\frac{1}{\bar{\xi}}.$$

三、估计量的评价标准

从之前的例子可以看到, 对于同一个参数, 用不同的估计方法求出的估计量可能不相同, 如例6.1.3(P14)和例6.2.3(P31)， 而且, 很明显, 原则上任何统计量都可以作为未知参数的估计量.

问题

(1)对于同一个参数究竟采用哪一个估计量好?

(2)评价估计量的标准是什么?

下面介绍几个常用标准.

1. 相合性

若$\hat{\theta}_n=\hat{\theta}_n(X_1,X_2,\cdots,X_n)$为参数$\theta$的估计量，若对于任意$\theta\in\Theta$，当$n\to\infty$时，$\hat{\theta}_n(X_1,X_2,\cdots,X_n)$依概率收敛于$\theta$，则称$\hat{\theta}_n$为$\theta$的相合估计量.

$$\lim_{n\to\infty}P\{|\hat{\theta}_n-\theta|<\varepsilon\}=1$$

注：矩估计量具有相合性

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{P} \mu_k, \quad k = 1, 2, \cdots;$$

由第四章关于依概率收敛的序列的性质知

$$\boxed{g(A_1, A_2, \cdots, A_k) \xrightarrow{P} g(\mu_1, \mu_2, \cdots, \mu_k),}$$

其中 $g(x_1, x_2, \cdots, x_n)$ 是有理函数，且 $g(\mu_1, \mu_2, \cdots, \mu_n) \neq \pm\infty$.

例6.2.6

设总体 X 服从参数为 $1/\theta$ 的指数分布, 概率密度

$$f(x;\theta)=\begin{cases}\dfrac{1}{\theta}\mathrm{e}^{-\frac{x}{\theta}}, & x>0,\\ 0, & \text{其他}.\end{cases}$$

其中参数 $\theta>0$, 又设 $X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本, 试证 $\overline{X}$ 是 θ 的相合估计量，而 $nZ=n[\min(X_1,X_2,\cdots,X_n)]$ 不是 θ 的相合估计量.

证明

因$E(X)=\theta<\infty$,由辛钦大数定律知,$\bar{X}$是相合估计量.

而 $Z=\min(X_1,X_2,\cdots,X_n)$ 服从参数为$\dfrac{n}{\theta}$的指数分布,

$$F_{\min}(x;\theta)=\begin{cases}1-\mathrm{e}^{-\frac{nx}{\theta}}, & x>0,\\ 0, & \text{其他}.\end{cases}$$

对$\forall\varepsilon>0$,

$$P\left(nZ-\theta>\varepsilon\right)=P\left(Z>\frac{\theta+\varepsilon}{n}\right)=e^{-\frac{\theta+\varepsilon}{\theta}}\not\to 0,\ as\ n\to\infty.$$

故nZ不是θ的相合估计量.

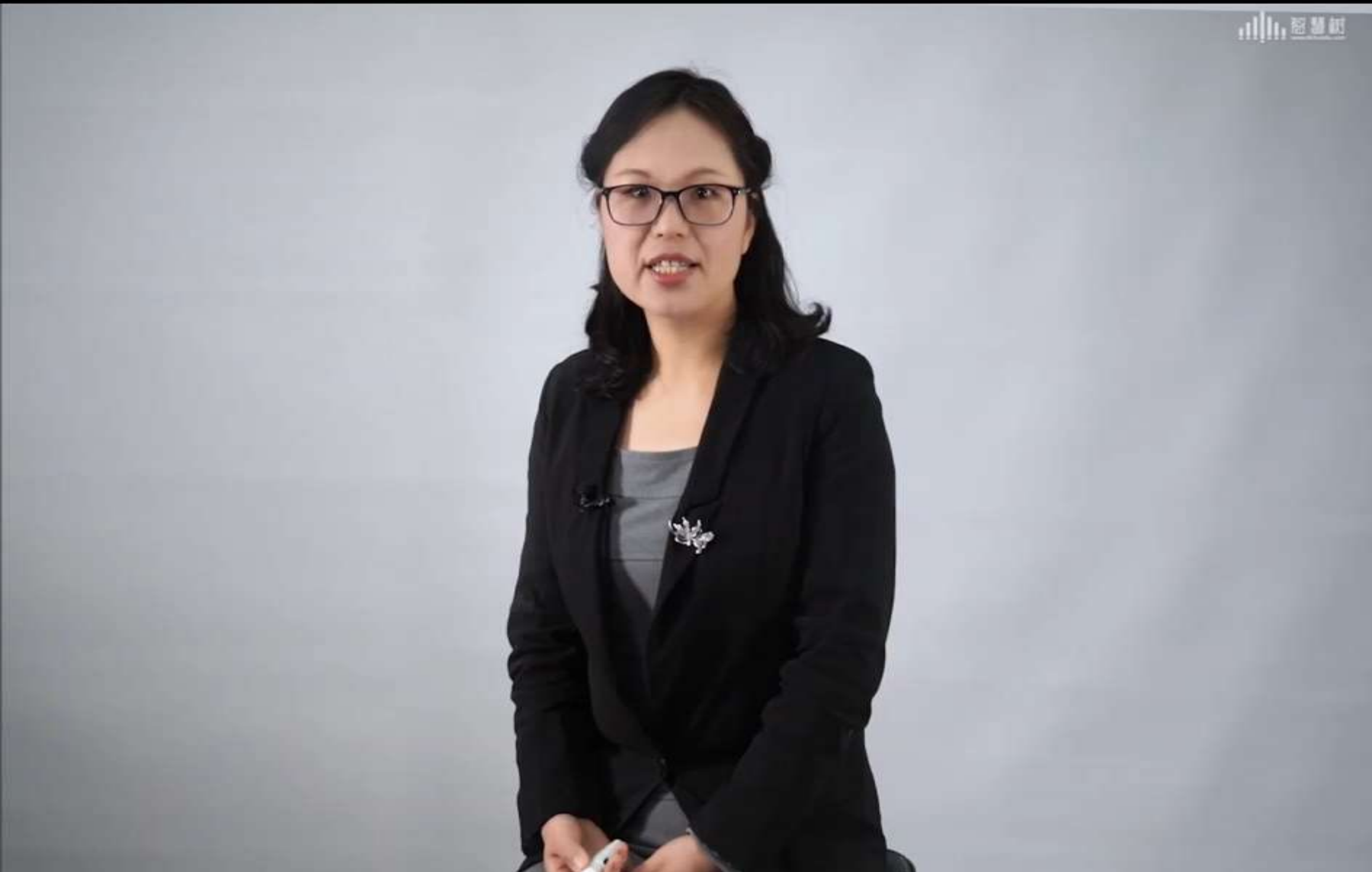
定理1

设$\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$是$\theta$的一个估计量，

若

$$\lim_{n\to\infty} E(\hat{\theta}_n) = \theta \qquad \lim_{n\to\infty} Var(\hat{\theta}_n) = 0$$

则$\hat{\theta}_n$是θ的相合估计。



2. 无偏性

若 $X_1, X_2, \cdots, X_n$ 为总体 X 的一个样本，$\theta \in \Theta$ 是包含在总体 X 的分布中的待估参数，(Θ 是 θ 的取值范围)

若估计量 $\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$ 的数学期望 $E(\hat{\theta})$ 存在，且对于任意 $\theta \in \Theta$ 有 $\boxed{E(\hat{\theta}) = \theta,}$ 则称 $\hat{\theta}$ 是 θ 的无偏估计量.

无偏估计的实际意义：无系统误差.

设总体X（不管服从什么分布，只要均值和方差存在）的均值为 μ ，方差为 σ^2 , $X_1, X_2, \cdots, X_n$ 是来自总体X的样本, $\bar{X}$, S_n^2 分别是样本均值和样本方差，则有

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

$$E(S_n^2) = \frac{n-1}{n}\sigma^2$$

结论

对于均值 μ , 方差 $\sigma^2>0$ 都存在的总体, 若 μ, σ^2 均为未知, 则

$E(\bar{X})=\mu$, $\bar{X}$是μ 的无偏估计量;

$E(S_n^2)\neq\sigma^2$, $S_n^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$ 是σ^2 的有偏估计量;

$E(S_n^{*2})=\sigma^2$, $S_n^{*2}=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ 是σ^2 的无偏估计量;

例6.2.7

设$X_1, X_2, \cdots, X_n$是来自总体X的样本,作为总体均值μ的估计量有

$$T_1 = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad T_2 = X_1, \quad T_3 = \sum_{i=1}^{n} a_i X_i$$

其中$a_i > 0 (i = 1,2 \cdots n)$,且$\sum_{i=1}^{n} a_i = 1$

试证T_1, T_2, T_3都是μ的无偏估计量;

解 $E(T_1) = E(\overline{X}) = \mu \qquad E(T_2) = E(X_1) = \mu$

$$E(T_3) = E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i) = \mu\left(\sum_{i=1}^{n} a_i\right) = \mu$$

由上例可知,一个参数可以有不同的无偏估计量.

例6.2.8 设 $\hat{\theta}$是参数 θ 的无偏估计，且有 $D(\hat{\theta})>0$ ，试证 $\tilde{\theta}=(\hat{\theta})^2$不是 θ^2的无偏估计.

例6.2.9 设总体服从参数为λ 的泊松分布,$X_1,\cdots X_n$ 是一简单随机样本，$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i, S_n^{*2}=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$ 。试证：$\frac{1}{2}\left(\overline{X}+S_n^{*2}\right)$是 λ 的无偏估计.

例6.2.10 设总体 X 在 $[0,\theta]$上服从均匀分布,参数$\theta>0$,
$X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本，试证明$\hat{\theta}_1=2\bar{X}$
和$\hat{\theta}_2=\dfrac{n+1}{n}\max(X_1,X_2,\cdots,X_n)$都是$\theta$的无偏估计.

证　因为　$E(2\bar{X})=2E(\bar{X})=2E(X)=2\times\dfrac{\theta}{2}=\theta,$

所以$\hat{\theta}_1=2\bar{X}$ 是 θ 的无偏估计量.

因　$X_{(n)}=\max(X_1,X_2,\cdots,X_n)$的概率密度

$$f(x)=\begin{cases}\dfrac{nx^{n-1}}{\theta^n}, & 0\le x\le\theta,\\ 0, & \text{其他}\end{cases}$$

所以 $E(X_{(n)}) = \int_0^{\theta} x \cdot \frac{nx^{n-1}}{\theta^n} \mathrm{d}x = \frac{n}{n+1}\theta,$

故有 $E\left(\frac{n+1}{n} X_{(n)}\right) = \theta,$

故 $\hat{\theta}_2 = \frac{n+1}{n} \max(X_1, X_2, \cdots, X_n)$ 也是 θ 的无偏估计量.

例6.2.6（续）

设总体 X 服从参数为 $1/\theta$ 的指数分布，概率密度

$$f(x;\theta)=\begin{cases}\dfrac{1}{\theta}\mathrm{e}^{-\frac{x}{\theta}}, & x>0,\\ 0, & \text{其他}.\end{cases}$$

其中参数 $\theta>0$，又设 $X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本，试证 $\overline{X}$ 和 $nZ=n[\min(X_1,X_2,\cdots,X_n)]$ 都是 θ 的无偏估计.

证明 因为 $E(\overline{X})=E(X)=\theta,$

所以 $\overline{X}$ 是 θ 的无偏估计量.

而 $Z = min(X_1, X_2, \cdots, X_n)$ 服从参数为 $\dfrac{n}{\theta}$ 的指数分布,

概率密度 $f_{\min}(x;\theta) = \begin{cases} \dfrac{n}{\theta} e^{-\frac{nx}{\theta}}, & x > 0, \\ 0, & 其他. \end{cases}$

故知 $E(Z) = \dfrac{\theta}{n}$, $E(nZ) = \theta$,

所以 nZ 也是 θ 的无偏估计量.

3. 有效性

比较参数θ的两个无偏估计量$\hat{\theta}_1$和$\hat{\theta}_2$，如果在样本容量n相同的情况下，$\hat{\theta}_1$的观察值在真值θ的附近较$\hat{\theta}_2$更密集，则认为$\hat{\theta}_1$较$\hat{\theta}_2$有效.

设$\hat{\theta}_1=\hat{\theta}_1(X_1,X_2,\cdots,X_n)$与$\hat{\theta}_2=\hat{\theta}_2(X_1,X_2,\cdots,X_n)$都是$\theta$的无偏估计量，若有 $D(\hat{\theta}_1)\le D(\hat{\theta}_2)$，

则称$\hat{\theta}_1$较$\hat{\theta}_2$有效.

例6.2.7(续)

设总体X的方差$D(X)$存在，试问T_1,T_2，T_3哪个更有效？

$$D(T_1)=D(\overline{X})=\frac{1}{n}D(X)=\frac{\sigma^2}{n}\qquad D(T_2)=D(X_1)=\sigma^2$$

$$D(T_3)=D\left(\sum_{i=1}^{n}a_iX_i\right)=\sum_{i=1}^{n}a_i^2D(X_i)=\sigma^2\left(\sum_{i=1}^{n}a_i^2\right)$$

注意 $\displaystyle\sum_{i=1}^{n}a_i^2\ge\frac{1}{n}$

$$\left(\sum_{i=1}^{n}x_iy_i\right)^2\le\left(\sum_{i=1}^{n}x_i^2\right)\left(\sum_{i=1}^{n}y_i^2\right).$$

所以$T_1=\overline{X}$是三个无偏估计量中最有效的估计量

例6. 2. 10　(续)

设总体 X 在 $[0,\theta]$上服从均匀分布,参数$\theta>0$, $X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本，试证明$\hat{\theta}_1=2\bar{X}$ 和$\hat{\theta}_2=\dfrac{n+1}{n}\max(X_1,X_2,\cdots,X_n)$都是$\theta$的无偏估计.

现证当$n\geq 2$时，$\hat{\theta}_2$较$\hat{\theta}_1$有效.

证明　由于 $D(\hat{\theta}_1)=D(2\bar{X})=4D(\bar{X})=\dfrac{4}{n}D(X)=\dfrac{\theta^2}{3n}$,

$$D(\hat{\theta}_2)=D\left(\frac{n+1}{n}X_{(n)}\right)=\left(\frac{n+1}{n}\right)^2D\left(X_{(n)}\right),$$

又因为 $E(X_{(n)})=\dfrac{n}{n+1}\theta,$

$$E(X_{(n)}{}^2)=\int_0^\theta \frac{n}{\theta^n}x^{n+1}\mathrm{d}x=\frac{n}{n+2}\theta^2,$$

$$D(X_{(n)})=E(X_{(n)}{}^2)-[E(X_{(n)})]^2$$

$$=\frac{n}{n+2}\theta^2-\left(\frac{n}{n+1}\theta\right)^2=\frac{n}{(n+1)^2(n+2)}\theta^2,$$

故 $D(\hat{\theta}_2)=\left(\dfrac{n+1}{n}\right)^2 D\left(X_{(n)}\right)=\dfrac{1}{n(n+2)}\theta^2<D(\hat{\theta}_1)=\dfrac{\theta^2}{3n},$

当$n\geq 2,$ 所以 $D(\hat{\theta}_2)<D(\hat{\theta}_1),$ $\hat{\theta}_2$ 较 $\hat{\theta}_1$ 有效.

例6.2.6（续2）

设总体 X 服从参数为 $1/\theta$ 的指数分布, 概率密度

$$f(x;\theta)=\begin{cases}\dfrac{1}{\theta}\mathrm{e}^{-\frac{x}{\theta}}, & x>0,\\ 0, & \text{其他}.\end{cases}$$

其中参数 $\theta>0$, 又设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本.

试证当 $n>1$ 时, θ 的无偏估计量 $\overline{X}$ 较 nZ 有效.

证明 由于 $D(X)=\theta^2$, 故有 $D(\overline{X})=\dfrac{\theta^2}{n}$,

而 $Z=\min(X_1,X_2,\cdots,X_n)$ 服从参数为 $\dfrac{n}{\theta}$ 的指数分布,

$$\text{概率密度 } f_{\min}(x;\theta)=\begin{cases}\dfrac{n}{\theta}e^{-\frac{nx}{\theta}}, & x>0,\\ 0, & \text{其他}.\end{cases}$$

又因为 $D(Z)=\dfrac{\theta^2}{n^2}$, 故有 $D(nZ)=\theta^2$,

当 $n>1$ 时, $D(nZ)>D(\overline{X})$,

故 θ 的无偏估计量 $\overline{X}$ 较 nZ 有效.

定理

极大似然估计的渐近正态性 设随机变量ξ 具有密度函数$f(x;\theta)$, 未知参数$\theta \in \Theta$,这里Θ 是一个非退化的开区间,并假定

(1) 对任何一个$\theta \in \Theta$ 和任一数x, 偏导数$\frac{\partial \ln f}{\partial \theta}, \frac{\partial^2 \ln f}{\partial \theta^2}, \frac{\partial^3 \ln f}{\partial \theta^3}$存在;

(2) 对Θ 中每一个θ,不等式

$$\left|\frac{\partial f}{\partial \theta}\right| < F_1(x), \left|\frac{\partial^2 f}{\partial \theta^2}\right| < F_2(x), \left|\frac{\partial^3 f}{\partial \theta^3}\right| < F_3(x)$$

成立,其中函数$F_1(x)$ 与$F_2(x)$ 在整个数轴$(-\infty, +\infty)$ 上可积,

而函数$F_3(x)$ 满足不等式

$$\int_{-\infty}^{+\infty} F_3(x)f(x;\theta)\mathrm{d}x < M$$

其中M 与θ 无关;

(3) 对Θ 中每一个θ 有,

$$0 < E\left(\frac{\partial \ln f}{\partial \theta}\right)^2 = \int_{-\infty}^{+\infty} \left(\frac{\partial \ln f}{\partial \theta}\right)^2 f(x;\theta)\mathrm{d}x < +\infty$$

于是若分布参数θ 的未知真值θ 是Θ 的一个内点,则方程

$$\frac{\partial \ln L}{\partial \theta} = 0$$

有一个解$\hat{\theta}_n = \hat{\theta}(\xi_1, \xi_2, \cdots, \xi_n)$ 存在，当$n \to \infty$ 时，

$$\hat{\theta}(\xi_1, \xi_2, \cdots, \xi_n) \xrightarrow{P} \theta,$$

且$\hat{\theta}_n$ 渐近地服从正态分布

$$\hat{\theta}_n \sim AN\left(\theta, \left[nE\left(\frac{\partial \ln f}{\partial \theta}\right)^2\right]^{-1}\right).$$

作业：

6. 2，6.5，6.7

6.8，6.21, 6.28，6.33，6.34，6.37

6.8，6.21，6.27，6.32，6.33，6.36（新版）