第六章　参数估计

实际工作中，对许多随机现象往往是知道其分布类型，但未知其中某些参数。如某全国统一考试的分数近似服从正态分布 $N(\mu,\sigma^2)$，但参数 μ,σ^2 未知；又如一场足球比赛进球数近似服从泊松分布 $P(\lambda)$，但参数 λ 未知。因此需根据样本对参数作出估计，分点估计（point estimation）和区间估计（interval estimation）两类。

§6.1　点估计的几种方法

设 θ 为总体 X 的未知参数，建立一个统计量 $\hat{\theta}$ 作为参数 θ 的估计量，称 $\hat{\theta}$ 为 θ 的点估计。常用的点估计法有矩估计和最大似然估计。

6.1.1　替换原理和矩估计

基本原理就是替换原理，用样本矩及其函数替换总体矩及其相应函数。

用样本原点矩 $A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$ 估计总体原点矩 $\mu_k=E(X^k)$；用样本中心矩 $B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k$ 估计总体中心矩 $\nu_k=E[X-E(X)]^k$。特别是用样本均值 $\bar{X}$ 估计总体期望 $E(X)$，即 $E\hat{(X)}=\bar{X}$；用未修正样本方差 $S^{*2}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$ 或样本方差 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ 估计总体方差 $\mathrm{Var}(X)$，即 $\mathrm{Var}\hat{(X)}=S^{*2}$ 或 S^2。

根据替换原理求总体参数的点估计，称为矩估计（Moment Estimation）。

对于具体问题，一般尽可能用低阶矩进行替换。首先计算总体期望 $E(X)$（一阶矩）。若 $E(X)$ 为总体参数 θ 的函数，即 $E(X)=g(\theta)$，反解出 $\theta=h[E(X)]$，用 $\bar{X}$ 替换 $E(X)$，可得矩估计 $\hat{\theta}=h(\bar{X})$；若 $E(X)$ 与总体参数 θ 无关或无法反解，再计算总体方差 $\mathrm{Var}(X)$（二阶矩），以此类推。

例　设总体 X 服从泊松分布 $P(\lambda)$，$X_1,X_2,\cdots,X_n$ 为样本，求 λ 的矩估计。

解：因 $E(X)=\lambda$，用 $\bar{X}$ 替换 $E(X)$，故 λ 的矩估计 $\hat{\lambda}=\bar{X}$。

例　设总体 X 服从指数分布 $Exp(\lambda)$，$X_1,X_2,\cdots,X_n$ 为样本，求 λ 的矩估计。

解：因 $E(X)=\frac{1}{\lambda}$，即 $\lambda=\frac{1}{E(X)}$，用 $\bar{X}$ 替换 $E(X)$，故 λ 的矩估计 $\hat{\lambda}=\frac{1}{\bar{X}}$。

例　设总体 X 的密度函数为

$$p(x)=\begin{cases}\theta x^{\theta-1}, & 0<x<1;\\ 0, & \text{其他}.\end{cases}$$

$X_1,X_2,\cdots,X_n$ 为样本，求 θ 的矩估计。

解：因

$$E(X)=\int_{-\infty}^{+\infty}xp(x)dx=\int_0^1 x\cdot\theta x^{\theta-1}dx=\int_0^1\theta x^{\theta}dx=\frac{\theta}{\theta+1}x^{\theta+1}\Big|_0^1=\frac{\theta}{\theta+1},$$

即$\theta=\dfrac{E(X)}{1-E(X)}$，用$\bar{X}$替换$E(X)$，故$\theta$的矩估计$\hat{\theta}=\dfrac{\bar{X}}{1-\bar{X}}$。

例　设总体X服从均匀分布$U(a,b)$，$X_1,X_2,\cdots,X_n$为样本，求a,b的矩估计。

解：因$E(X)=\dfrac{a+b}{2}$，$\mathrm{Var}(X)=\dfrac{(b-a)^2}{12}$，可得

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

用$\bar{X}$替换$E(X)$，未修正样本方差S^{*2}替换$\mathrm{Var}(X)$，故a,b的矩估计$\hat{a}=\bar{X}-\sqrt{3}S^*$，$\hat{b}=\bar{X}+\sqrt{3}S^*$。

6.1.2　最大似然估计

基本原理：样本观测值可能在某参数取各种不同值下发生，使该样本观测值出现的概率最大的参数值作为该参数的估计值。

将样本联合质量函数或联合密度函数$p(x_1,x_2,\cdots,x_n;\theta)=p(x_1;\theta)p(x_2;\theta)\cdots p(x_n;\theta)$看作关于参数$\theta$的函数$L(\theta)$，称为似然函数。

定义　若似然函数$L(\theta)$在$\hat{\theta}$达到最大值，则称$L(\theta)$的最大值点$\hat{\theta}$是参数θ的最大似然估计（Maximum Likelihood Estimation，简称 MLE）。

为了求似然函数$L(\theta)$的最大值点$\hat{\theta}$，可以令$\dfrac{dL(\theta)}{d\theta}=0$，称为似然方程，再解得$\hat{\theta}$。但由于$L(\theta)$是乘积式，为了求导方便，通常取对数$\ln L(\theta)$，且显然$L(\theta)$与其对数$\ln L(\theta)$同时达到最大值，将似然方程改为$\dfrac{d\ln L(\theta)}{d\theta}=0$。

求最大似然估计的一般步骤：

（1）建立似然函数$L(\theta)=p(x_1,x_2,\cdots,x_n;\theta)$。

（2）取对数$\ln L(\theta)$，求导并令为 0，得似然方程$\dfrac{d\ln L(\theta)}{d\theta}=0$。

（3）解得驻点$\theta=\hat{\theta}(x_1,x_2,\cdots,x_n)$，并判断是否最大值点。

（4）得出最大似然估计$\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$。

例　已知总体X服从指数分布$Exp(\lambda)$，$X_1,X_2,\cdots,X_n$为样本，求λ的最大似然估计。

解：因指数分布$Exp(\lambda)$的密度函数为

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

则似然函数为

$$L(\lambda)=p(x_1,x_2,\cdots,x_n;\lambda)=\prod_{i=1}^{n}\lambda e^{-\lambda x_i}I_{x_i>0}=\lambda^n e^{-\lambda\sum\limits_{i=1}^{n}x_i}I_{x_1,x_2,\cdots,x_n>0}。$$

当$x_1,x_2,\cdots,x_n>0$时，有

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

令

$$\frac{d\ln L(\lambda)}{d\lambda}=\frac{n}{\lambda}-\sum_{i=1}^n x_i=0\text{，}$$

解得驻点

$$\lambda=\frac{n}{\sum_{i=1}^n x_i}=\frac{1}{\bar{x}}\text{ 。}$$

又因

$$\frac{d^2\ln L(\lambda)}{d\lambda^2}=-\frac{n}{\lambda^2}<0\text{，}$$

故 $\lambda=\dfrac{1}{\bar{x}}$ 是最大值点，λ 的最大似然估计为 $\hat{\lambda}=\dfrac{1}{\bar{X}}$ 。

例 已知总体 X 服从泊松分布 $P(\lambda)$， $X_1, X_2, \cdots, X_n$ 为样本，求 λ 的最大似然估计。

解：因泊松分布 $P(\lambda)$ 的质量函数为

$$p(x;\lambda)=\frac{\lambda^x}{x!}e^{-\lambda},\quad x=0,1,2,\cdots\text{，}$$

则似然函数为

$$L(\lambda)=\prod_{i=1}^n\frac{\lambda^{x_i}}{x_i!}e^{-\lambda}=\frac{\lambda^{\sum_{i=1}^n x_i}}{x_1!x_2!\cdots x_n!}e^{-n\lambda},\quad x_1,x_2,\cdots,x_n=0,1,2,\cdots\text{ 。}$$

有

$$\ln L(\lambda)=\ln\lambda^{\sum_{i=1}^n x_i}+\ln e^{-n\lambda}-\ln(x_1!x_2!\cdots x_n!)=\ln\lambda\cdot\sum_{i=1}^n x_i-n\lambda-\sum_{i=1}^n\ln(x_i!)\text{ 。}$$

令

$$\frac{d\ln L(\lambda)}{d\lambda}=\frac{1}{\lambda}\sum_{i=1}^n x_i-n=0\text{，}$$

解得驻点

$$\lambda=\frac{1}{n}\sum_{i=1}^n x_i=\bar{x}\text{ 。}$$

又因

$$\frac{d^2\ln L(\lambda)}{d\lambda^2}=-\frac{1}{\lambda^2}\sum_{i=1}^n x_i<0\text{，}$$

故 $\lambda=\bar{x}$ 是最大值点，λ 的最大似然估计为 $\hat{\lambda}=\bar{X}$ 。

例 设总体 X 服从正态分布 $N(\mu,\sigma^2)$， $X_1, X_2, \cdots, X_n$ 为样本，求 μ,σ^2 的最大似然估计。

解：因正态分布 $N(\mu,\sigma^2)$ 的密度函数为

$$p(x;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\text{，}$$

则似然函数为

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

有

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

令

$$\begin{cases}\dfrac{\partial\ln L(\mu,\sigma^2)}{\partial\mu}=-\dfrac{1}{2\sigma^2}\sum_{i=1}^{n}2(x_i-\mu)\cdot(-1)=\dfrac{1}{\sigma^2}\left(\sum_{i=1}^{n}x_i-n\mu\right)=0;\\ \dfrac{\partial\ln L(\mu,\sigma^2)}{\partial(\sigma^2)}=-\dfrac{n}{2}\cdot\dfrac{1}{\sigma^2}+\dfrac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2=-\dfrac{n}{2\sigma^2}+\dfrac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2=0.\end{cases}$$

解得

$$\mu=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}\text{，}\quad\sigma^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2=s*^2\text{。}$$

又因

$$A=\left.\frac{\partial^2\ln L(\mu,\sigma^2)}{\partial\mu^2}\right|_{\substack{\mu=\bar{x}\\ \sigma^2=s*^2}}=-\left.\frac{n}{\sigma^2}\right|_{\substack{\mu=\bar{x}\\ \sigma^2=s*^2}}=-\frac{n}{s*^2}\text{，}$$

$$B=\left.\frac{\partial^2\ln L(\mu,\sigma^2)}{\partial\mu\partial(\sigma^2)}\right|_{\substack{\mu=\bar{x}\\ \sigma^2=s*^2}}=-\frac{1}{\sigma^4}\left.\left(\sum_{i=1}^{n}x_i-n\mu\right)\right|_{\substack{\mu=\bar{x}\\ \sigma^2=s*^2}}=0\text{，}$$

$$C=\left.\frac{\partial^2\ln L(\mu,\sigma^2)}{\partial(\sigma^2)^2}\right|_{\substack{\mu=\bar{x}\\ \sigma^2=s*^2}}=\frac{n}{2\sigma^4}-\frac{1}{\sigma^6}\left.\sum_{i=1}^{n}(x_i-\mu)^2\right|_{\substack{\mu=\bar{x}\\ \sigma^2=s*^2}}=-\frac{n}{2s*^4}\text{，}$$

有

$$A=-\frac{n}{s*^2}<0\text{，}\quad B^2-AC=-\frac{n^2}{2s*^6}<0\text{，}$$

故 $(\mu,\sigma^2)=(\bar{x},s*^2)$ 是最大值点，μ,σ^2 的最大似然估计分别为 $\hat{\mu}=\bar{X}$ 、$\hat{\sigma}^2=S*^2$ 。

可见，若总体服从指数分布 $Exp(\lambda)$ 、泊松分布 $P(\lambda)$ 、正态分布 $N(\mu,\sigma^2)$ 等，则相应参数的矩估计与最大似然估计相同。但一般情形下，参数的矩估计与最大似然估计不一定相同。

例 设总体 X 的密度函数为

$$p(x;\theta)=\begin{cases}\theta x^{\theta-1}, & 0<x<1;\\ 0, & \text{其他.}\end{cases}$$

$X_1,X_2,\cdots,X_n$ 为样本，求 θ 的最大似然估计。

解： 似然函数为

$$L(\theta)=\prod_{i=1}^{n}\theta x_i^{\theta-1}I_{0<x_i<1}=\theta^n(x_1x_2\cdots x_n)^{\theta-1}I_{0<x_1,x_2,\cdots,x_n<1}\text{。}$$

当 $0<x_1,x_2,\cdots,x_n<1$ 时，有

$$\ln L(\theta)=\ln\theta^n+\ln(x_1x_2\cdots x_n)^{\theta-1}=n\ln\theta+(\theta-1)\sum_{i=1}^n\ln x_i \text{ 。}$$

令

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

解得驻点

$$\theta=-\frac{n}{\sum\limits_{i=1}^n\ln x_i}\text{ 。}$$

又因

$$\frac{d^2\ln L(\theta)}{d\theta^2}=-\frac{n}{\theta^2}<0\text{，}$$

故 $\theta=-\dfrac{n}{\sum\limits_{i=1}^n\ln x_i}$ 是最大值点，θ 的最大似然估计为 $\hat\theta=-\dfrac{n}{\sum\limits_{i=1}^n\ln X_i}$ 。

注：前面例子中，该分布参数 θ 的矩估计为 $\hat\theta=\dfrac{\bar X}{1-\bar X}$，其矩估计与最大似然估计不同。

例 已知总体 X 服从均匀分布 $U(0,\theta)$， $X_1, X_2,\cdots, X_n$ 为样本，求 θ 的矩估计与最大似然估计。

解：矩估计，因 $E(X)=\dfrac{\theta}{2}$，即 $\theta=2E(X)$，用 $\bar X$ 替换 $E(X)$，故 θ 的矩估计 $\hat\theta=2\bar X$ 。

最大似然估计，因均匀分布 $U(0,\theta)$ 的密度函数为

$$p(x;\theta)=\frac{1}{\theta}I_{0<x<\theta}\text{，}$$

则似然函数为

$$L(\theta)=\prod_{i=1}^n\frac{1}{\theta}I_{0<x_i<\theta}=\frac{1}{\theta^n}I_{0<x_1,\,x_2,\,\cdots,\,x_n<\theta}\text{ 。}$$

按常规方法，当 $0<x_1, x_2,\cdots, x_n<\theta$ 时，有

$$\ln L(\theta)=-\ln\theta^n=-n\ln\theta\text{ 。}$$

令

$$\frac{d\ln L(\theta)}{d\theta}=-\frac{n}{\theta}=0\text{，无解。}$$

此时应直接根据似然函数找出最大值点。因似然函数 $L(\theta)=\dfrac{1}{\theta^n}I_{0<x_1,\,x_2,\,\cdots,\,x_n<\theta}$， 参数 θ 越小， $\dfrac{1}{\theta^n}$ 越大，但只有当 $0<x_1, x_2,\cdots, x_n<\theta$ 时，才有 $L(\theta)>0$ 。可见要使得 $L(\theta)$ 达到最大，一方面 θ 应尽可能小，另一方面 θ 应大于 $x_1, x_2,\cdots, x_n$ 每一个值，即大于 $\max\{x_1, x_2,\cdots, x_n\}$ 。故 $\theta=\max\{x_1, x_2,\cdots, x_n\}=x_{(n)}$ 是最大值点，θ 的最大似然估计为 $\hat\theta=X_{(n)}$ 。

注：虽然这里本来应该是 $\theta>\max\{x_1, x_2,\cdots, x_n\}=x_{(n)}$， θ 不能取到 $x_{(n)}$，但由于有限个点上取值不相同的密度函数对应于同一个分布，因此可以将密度函数改为 x 可以取到 θ，这并不改变其分布。

§6.2　点估计的评价标准

不同的估计方法，可得不同的估计量，需比较其好坏。

6.2.1　　相合性

对于未知参数θ的估计量$\hat{\theta}$，由于$\hat{\theta}$是统计量（随机变量），而θ是一个数值，希望$\hat{\theta}$“稳定”在参数θ的附近，即$\hat{\theta}$能依概率收敛于参数θ。

定义　设θ为总体X的未知参数，$X_1, X_2, \cdots, X_n$为样本，$\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \cdots, X_n)$为参数$\theta$的估计量。若对任意的$\varepsilon > 0$，都有

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

则称$\hat{\theta}_n$是参数θ的相合估计。

相合性就是要求估计量依概率收敛于未知参数。

由辛钦大数定律知样本均值$\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$依概率收敛于总体期望$E(X) = \mu$，即样本均值$\overline{X}$是总体期望$\mu$的相合估计。

定理　设$\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \cdots, X_n)$为参数$\theta$的估计量，若

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

则$\hat{\theta}_n$是参数θ的相合估计。

证明：设$Y = \hat{\theta}_n - \theta$，首先证明不等式

$$P\{|Y| \ge \varepsilon\} \le \frac{E(Y^2)}{\varepsilon^2}。$$

以连续情形为例证明，设Y的密度函数为$p(y)$，有

$$P\{|Y| \ge \varepsilon\} = \int_{|y|\ge\varepsilon} p(y)dy \text{ 且 } \frac{E(Y^2)}{\varepsilon^2} = \int_{-\infty}^{+\infty} \frac{y^2}{\varepsilon^2} p(y)dy,$$

故

$$P\{|Y| \ge \varepsilon\} = \int_{|y|\ge\varepsilon} p(y)dy \le \int_{|y|\ge\varepsilon} \frac{y^2}{\varepsilon^2} p(y)dy \le \int_{-\infty}^{+\infty} \frac{y^2}{\varepsilon^2} p(y)dy = \frac{E(Y^2)}{\varepsilon^2}。$$

因$\lim\limits_{n\to+\infty} E(\hat{\theta}_n) = \theta$，$\lim\limits_{n\to+\infty} \mathrm{Var}(\hat{\theta}_n) = 0$，有

$$\lim_{n\to+\infty} E[(\hat{\theta}_n - \theta)^2] = \lim_{n\to+\infty} \{\mathrm{Var}(\hat{\theta}_n - \theta) + [E(\hat{\theta}_n - \theta)]^2\} = \lim_{n\to+\infty} \mathrm{Var}(\hat{\theta}_n) + \lim_{n\to+\infty} [E(\hat{\theta}_n) - \theta]^2 = 0,$$

故

$$\lim_{n\to+\infty} P\{|\hat{\theta}_n - \theta| \ge \varepsilon\} \le \lim_{n\to+\infty} \frac{E[(\hat{\theta}_n - \theta)^2]}{\varepsilon^2} = 0,$$

即$\hat{\theta}_n$是参数θ的相合估计。

注：若 $\lim\limits_{n\to+\infty} E[(\hat{\theta}_n-\theta)^2]=0$，称为 $\hat{\theta}_n$ 均方收敛于 θ。这里证明了均方收敛必然依概率收敛。但需注意的是依概率收敛不一定均方收敛。

例　设总体 X 服从正态分布 $N(\mu,\sigma^2)$，$X_1, X_2, \cdots, X_n$ 为样本。判断样本方差 $S^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2$ 与未修正样本方差 $S^{*2}=\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2$ 是否 σ^2 的相合估计。

解：因

$$\chi^2=\frac{(n-1)S^2}{\sigma^2}=\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2\sim\chi^2(n-1),$$

有 $E(\chi^2)=n-1$，$\mathrm{Var}(\chi^2)=2(n-1)$，则

$$E\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right]=(n-1)\sigma^2,\quad \mathrm{Var}\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right]=2(n-1)\sigma^4。$$

因 $S^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2$，有

$$E(S^2)=\frac{1}{n-1}\cdot(n-1)\sigma^2=\sigma^2,\quad \mathrm{Var}(S^2)=\frac{1}{(n-1)^2}\cdot 2(n-1)\sigma^4=\frac{2}{n-1}\sigma^4,$$

则

$$\lim_{n\to+\infty}E(S^2)=\lim_{n\to+\infty}\sigma^2=\sigma^2,\quad \lim_{n\to+\infty}\mathrm{Var}(S^2)=\lim_{n\to+\infty}\frac{2}{n-1}\sigma^4=0,$$

故样本方差 S^2 是 σ^2 的相合估计。

因 $S^{*2}=\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2$，有

$$E(S^{*2})=\frac{1}{n}\cdot(n-1)\sigma^2=\frac{n-1}{n}\sigma^2,\quad \mathrm{Var}(S^{*2})=\frac{1}{n^2}\cdot 2(n-1)\sigma^4=\frac{2(n-1)}{n^2}\sigma^4,$$

则

$$\lim_{n\to+\infty}E(S^{*2})=\lim_{n\to+\infty}\frac{n-1}{n}\sigma^2=\sigma^2,\quad \lim_{n\to+\infty}\mathrm{Var}(S^{*2})=\lim_{n\to+\infty}\frac{2(n-1)}{n^2}\sigma^4=0,$$

故未修正样本方差 S^{*2} 也是 σ^2 的相合估计。

一般地，设 $\hat{\theta}_n=\hat{\theta}_n(X_1, X_2, \cdots, X_n)$ 是参数 θ 的相合估计，且满足 $\lim\limits_{n\to+\infty}E(\hat{\theta}_n)=\theta$，$\lim\limits_{n\to+\infty}\mathrm{Var}(\hat{\theta}_n)=0$。若 $\{a_n\}$ 是任意一个极限等于 1 的数列，则

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

即 $a_n\hat{\theta}_n$ 也是参数 θ 的相合估计。

定理　若 $\hat{\theta}_{n1}, \hat{\theta}_{n2}, \cdots, \hat{\theta}_{nk}$ 分别是参数 $\theta_1, \theta_2, \cdots, \theta_k$ 的相合估计，$\eta=g(\theta_1, \theta_2, \cdots, \theta_k)$ 是 k 元连续函数，则 $\hat{\eta}_n=g(\hat{\theta}_{n1}, \hat{\theta}_{n2}, \cdots, \hat{\theta}_{nk})$ 是 η 的相合估计，即连续函数保持相合性。

根据辛钦大数定律，知样本矩 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ 依概率收敛于总体矩 $\mu_k = E(X^k)$，即样本矩是总体矩的相合估计，再根据连续函数保持相合性可知参数的矩估计都是其相合估计。

6.2.2　　无偏性

对于未知参数 θ 的估计量 $\hat{\theta}$，由于 $\hat{\theta}$ 是统计量（随机变量），而 θ 是一个数值，希望 $\hat{\theta}$ 以 θ 为中心。

定义　设 $\hat{\theta}$ 是未知参数 θ 的一个估计量。若 $E(\hat{\theta}) = \theta$，则称 $\hat{\theta}$ 是 θ 的一个无偏估计（Unbiased Estimation，简称 UE）。

如样本均值 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 作为总体期望 $E(X)$ 的估计量，有 $E(\bar{X}) = E(X)$，故 $\bar{X}$ 是 $E(X)$ 的无偏估计；样本方差 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ 作为总体方差 $\mathrm{Var}(X)$ 的估计量，有 $E(S^2) = \mathrm{Var}(X)$，故 S^2 是 $\mathrm{Var}(X)$ 的无偏估计。但因 $E(S^{*2}) = \frac{n-1}{n}\mathrm{Var}(X) \neq \mathrm{Var}(X)$，故未修正样本方差 $S^{*2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ 不是 $\mathrm{Var}(X)$ 的无偏估计。

注：根据替换原理从直观上应该用未修正样本方差 $S^{*2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ 估计总体方差 $\mathrm{Var}(X)$。但由于 S^{*2} 不是 $\mathrm{Var}(X)$ 的无偏估计，而是 $E\left(\frac{n}{n-1}S^{*2}\right) = \mathrm{Var}(X)$，因此取 $S^2 = \frac{n}{n-1}S^{*2}$ 作为 $\mathrm{Var}(X)$ 的无偏估计，这称之为对 S^{*2} 修偏。

一般地，对于总体未知参数 θ，根据某种方法，建立 θ 的估计量 $\hat{\theta}_n^* = \hat{\theta}_n^*(X_1, X_2, \cdots, X_n)$，判断 $\hat{\theta}_n^*$ 是否是 θ 的无偏估计。如果不满足无偏性，$E(\hat{\theta}_n^*) \neq \theta$，则根据 $E(\hat{\theta}_n^*)$ 的具体结果找出 $\hat{\theta}_n$ 使得 $E(\hat{\theta}_n) = \theta$，从而 $\hat{\theta}_n$ 是 θ 的无偏估计，这称之为修偏。

如 $E(\hat{\theta}_n^*) = a\theta + b \neq \theta$（$a \neq 0$），则有 $E\left(\frac{\hat{\theta}_n^* - b}{a}\right) = \theta$，这样修偏得到 $\hat{\theta}_n = \frac{\hat{\theta}_n^* - b}{a}$ 是 θ 的无偏估计。

例　设总体 X 服从均匀分布 $U(0,\theta)$，$X_1, X_2, \cdots, X_n$ 为样本，参数 θ 的矩估计 $\hat{\theta}_1^* = 2\bar{X}$ 与最大似然估计 $\hat{\theta}_2^* = X_{(n)}$ 是否参数 θ 的相合估计与无偏估计？如果不是无偏的，如何修偏使之成为 θ 的无偏估计？

解：矩估计 $\hat{\theta}_1^* = 2\bar{X}$，有

$$E(\hat{\theta}_1^*) = 2E(\bar{X}) = 2E(X) = \theta,\quad \mathrm{Var}(\hat{\theta}_1^*) = 4\mathrm{Var}(\bar{X}) = \frac{4}{n}\mathrm{Var}(X) = \frac{4}{n}\cdot\frac{\theta^2}{12} = \frac{1}{3n}\theta^2,$$

因

$$\lim_{n\to+\infty} E(\hat{\theta}_1^*) = \theta,\quad \lim_{n\to+\infty}\mathrm{Var}(\hat{\theta}_1^*) = \lim_{n\to+\infty}\frac{1}{3n}\theta^2 = 0,$$

故矩估计 $\hat{\theta}_1^* = 2\bar{X}$ 是参数 θ 的相合估计。又因 $E(\hat{\theta}_1^*) = \theta$，可知 $\hat{\theta}_1^*$ 也是参数 θ 的无偏估计。

最大似然估计 $\hat{\theta}_2^* = X_{(n)}$，因总体 X 的密度函数与分布函数分别为

$$p(x)=\frac{1}{\theta}I_{0<x<\theta}，\quad F(x)=\begin{cases}0, & x<0;\\ \dfrac{x}{\theta}, & 0\le x<\theta;\\ 1, & x\ge\theta.\end{cases}$$

则 $X_{(n)}$ 的密度函数为

$$p_n(x)=n[F(x)]^{n-1}p(x)=\frac{nx^{n-1}}{\theta^n}I_{0<x<\theta}，$$

可得

$$E(\hat{\theta}_2*)=E(X_{(n)})=\int_{-\infty}^{+\infty}xp_n(x)dx=\int_0^{\theta}\frac{n}{\theta^n}x^n dx=\frac{n}{\theta^n}\cdot\frac{x^{n+1}}{n+1}\bigg|_0^{\theta}=\frac{n}{n+1}\theta，$$

$$E(\hat{\theta}_2*^2)=E(X_{(n)}^2)=\int_{-\infty}^{+\infty}x^2p_n(x)dx=\int_0^{\theta}\frac{n}{\theta^n}x^{n+1}dx=\frac{n}{\theta^n}\cdot\frac{x^{n+2}}{n+2}\bigg|_0^{\theta}=\frac{n}{n+2}\theta^2，$$

$$\mathrm{Var}(\hat{\theta}_2*)=E(\hat{\theta}_2*^2)-[E(\hat{\theta}_2*)]^2=\frac{n}{n+2}\theta^2-\left(\frac{n}{n+1}\theta\right)^2=\frac{n}{(n+2)(n+1)^2}\theta^2。$$

因

$$\lim_{n\to+\infty}E(\hat{\theta}_2*)=\lim_{n\to+\infty}\frac{n}{n+1}\theta=\theta，\quad \lim_{n\to+\infty}\mathrm{Var}(\hat{\theta}_2*)=\lim_{n\to+\infty}\frac{n}{(n+2)(n+1)^2}\theta^2=0，$$

故最大似然估计 $\hat{\theta}_2*=X_{(n)}$ 是参数 θ 的相合估计。又因

$$E(\hat{\theta}_2*)=\frac{n}{n+1}\theta\ne\theta，\ 即 E\left(\frac{n+1}{n}\hat{\theta}_2*\right)=E\left(\frac{n+1}{n}X_{(n)}\right)=\theta，$$

可知 $\hat{\theta}_2*=X_{(n)}$ 不是参数 θ 的无偏估计，修偏成为 $\hat{\theta}_2=\dfrac{n+1}{n}X_{(n)}$ 才是 θ 的无偏估计。

6.2.3　有效性

若几个估计量都是无偏估计，就希望方差越小越好。

定义　设 $\hat{\theta}_1,\hat{\theta}_2$ 是未知参数 θ 的两个无偏估计，若 $\mathrm{Var}(\hat{\theta}_1)<\mathrm{Var}(\hat{\theta}_2)$，则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 更有效。

例　设总体 X 的期望为 μ，方差为 σ^2，X_1,X_2,X_3 为样本，且

$$\hat{\mu}_1=\frac{1}{3}X_1+\frac{1}{2}X_2+\frac{1}{6}X_3，\quad \hat{\mu}_2=\frac{1}{3}X_1+\frac{1}{3}X_2+\frac{1}{3}X_3，$$

$$\hat{\mu}_3=\frac{1}{4}X_1+\frac{1}{4}X_2+\frac{1}{4}X_3，\quad \hat{\mu}_4=2X_1+2X_2-3X_3，$$

判断 $\hat{\mu}_1,\hat{\mu}_2,\hat{\mu}_3,\hat{\mu}_4$ 中哪些是 μ 的无偏估计？在 μ 的无偏估计中哪一个最有效。

解：因

$$E(\hat{\mu}_1)=\frac{1}{3}E(X_1)+\frac{1}{2}E(X_2)+\frac{1}{6}E(X_3)=\frac{1}{3}\mu+\frac{1}{2}\mu+\frac{1}{6}\mu=\mu，$$

$$E(\hat{\mu}_2)=\frac{1}{3}\mu+\frac{1}{3}\mu+\frac{1}{3}\mu=\mu,$$

$$E(\hat{\mu}_3)=\frac{1}{4}\mu+\frac{1}{4}\mu+\frac{1}{4}\mu=\frac{3}{4}\mu\neq\mu,$$

$$E(\hat{\mu}_4)=2\mu+2\mu-3\mu=\mu,$$

故 $\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_4$ 是 μ 的无偏估计， $\hat{\mu}_3$ 不是 μ 的无偏估计。

因

$$\mathrm{Var}(\hat{\mu}_1)=\frac{1}{9}\mathrm{Var}(X_1)+\frac{1}{4}\mathrm{Var}(X_2)+\frac{1}{36}\mathrm{Var}(X_3)=\frac{1}{9}\sigma^2+\frac{1}{4}\sigma^2+\frac{1}{36}\sigma^2=\frac{14}{36}\sigma^2,$$

$$\mathrm{Var}(\hat{\mu}_2)=\frac{1}{9}\sigma^2+\frac{1}{9}\sigma^2+\frac{1}{9}\sigma^2=\frac{1}{3}\sigma^2,$$

$$\mathrm{Var}(\hat{\mu}_4)=4\sigma^2+4\sigma^2+9\sigma^2=17\sigma^2,$$

有

$$\mathrm{Var}(\hat{\mu}_2)<\mathrm{Var}(\hat{\mu}_1)<\mathrm{Var}(\hat{\mu}_4)$$

故 $\hat{\mu}_2$ 最有效， $\hat{\mu}_1$ 其次， $\hat{\mu}_4$ 最差。

注：有效性必须以无偏性为前提。此例中尽管 $\hat{\mu}_3$ 的方差最小，但它不是无偏估计，不考虑其有效性。

例 设总体 X 期望为 μ，方差为 σ^2， $X_1, X_2, \cdots, X_n$ 为样本。样本线性组合 $\hat{\mu}=a_1X_1+a_2X_2+\cdots+a_nX_n$ 称为期望 μ 的线性估计。当 $a_1, a_2, \cdots, a_n$ 满足什么条件时，线性估计 $\hat{\mu}$ 是期望 μ 的无偏估计？在什么情况下线性估计 $\hat{\mu}$ 最有效？

解：因

$$E(\hat{\mu})=a_1E(X_1)+a_2E(X_2)+\cdots+a_nE(X_n)=(a_1+a_2+\cdots+a_n)\mu,$$

故当 $a_1+a_2+\cdots+a_n=1$ 时， $E(\hat{\mu})=\mu$，线性估计 $\hat{\mu}$ 是期望 μ 的无偏估计。又因

$$\mathrm{Var}(\hat{\mu})=a_1^2\,\mathrm{Var}(X_1)+a_2^2\,\mathrm{Var}(X_2)+\cdots+a_n^2\,\mathrm{Var}(X_n)=(a_1^2+a_2^2+\cdots+a_n^2)\sigma^2,$$

在 $a_1+a_2+\cdots+a_n=1$ 的条件下，使得 $a_1^2+a_2^2+\cdots+a_n^2$ 达到最小时，线性估计 $\hat{\mu}$ 最有效。

设拉格朗日函数

$$L(a_1, a_2, \cdots, a_n, \lambda)=a_1^2+a_2^2+\cdots+a_n^2+\lambda(a_1+a_2+\cdots+a_n-1),$$

令

$$\begin{cases}\dfrac{\partial L}{\partial a_i}=2a_i+\lambda=0, \quad i=1, 2, \cdots, n;\\[2mm] \dfrac{\partial L}{\partial \lambda}=a_1+a_2+\cdots+a_n-1=0.\end{cases}$$

可得

$$a_1 = a_2 = \cdots = a_n = \frac{1}{n},$$

可见样本均值 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 是总体期望 μ 最有效的**线性**无偏估计。

那么，样本均值 $\bar{X}$ 是否一定是总体期望 μ 最有效的无偏估计？也就是说在总体期望 μ 的无偏估计中是否一定不能找到比样本均值 $\bar{X}$ 更有效的估计量？答案是不一定。

例　设总体 X 服从均匀分布 $U(0,\theta)$，$X_1, X_2, \cdots, X_n$ 为样本，$\hat{\theta}_1 = 2\bar{X}$ 与 $\hat{\theta}_2 = \frac{n+1}{n} X_{(n)}$ 都是参数 θ 的无偏估计，比较它们的有效性。

解：因

$$\mathrm{Var}(\hat{\theta}_1) = 4\,\mathrm{Var}(\bar{X}) = \frac{4}{n}\mathrm{Var}(X) = \frac{4}{n}\cdot\frac{\theta^2}{12} = \frac{1}{3n}\theta^2,$$

$$\mathrm{Var}(\hat{\theta}_2) = \frac{(n+1)^2}{n^2}\mathrm{Var}(X_{(n)}) = \frac{(n+1)^2}{n^2}\cdot\frac{n}{(n+2)(n+1)^2}\theta^2 = \frac{1}{n(n+2)}\theta^2,$$

当 $n>1$ 时，$\mathrm{Var}(\hat{\theta}_2) < \mathrm{Var}(\hat{\theta}_1)$，即 $\hat{\theta}_2 = \frac{n+1}{n} X_{(n)}$ 比 $\hat{\theta}_1 = 2\bar{X}$ 作为参数 θ 的无偏估计更有效。

注：因 $E(X) = \frac{\theta}{2}$，可知 $\bar{X}$ 与 $\frac{n+1}{2n} X_{(n)}$ 都是总体期望 $E(X)$ 的无偏估计，而 $\frac{n+1}{2n} X_{(n)}$ 比 $\bar{X}$ 更有效。

§6.3 均方误差

通常情况下，评价一个估计量的好坏，首先判断是否无偏估计，在无偏估计的情况下再进一步判断有效性。但有时也可以综合考虑无偏性和有效性，而考虑其均方误差。

如某参数θ的真值为 50，现有三个估计量$\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3$，且$\hat{\theta}_1$以 1/3 的概率分别取值 40, 50, 60；$\hat{\theta}_2$以 1/3 的概率分别取值 49.5, 50.5, 51.5；$\hat{\theta}_3$以 1/3 的概率分别取值 54.9, 55, 55.1。从无偏性考虑，只有$\hat{\theta}_1$才满足，但显然其方差太大，有效性太差；而$\hat{\theta}_2$虽然不是θ的无偏估计，但其偏差不大，且方差较小；$\hat{\theta}_3$的方差是最小的，但其偏差又太大。综合考虑$\hat{\theta}_2$应该是最好的一个估计量。

对于参数θ的一个估计量$\hat{\theta}$，无论其是否无偏估计，考虑$\hat{\theta}$的估计值与参数θ的真值之间距离平方的期望值$E(\hat{\theta}-\theta)^2$，称为均方误差（Mean Square Error），记为$\mathrm{MSE}_\theta(\hat{\theta})$，希望均方误差越小越好。

均方误差的计算公式：

$$\mathrm{MSE}_\theta(\hat{\theta}) = E(\hat{\theta}-\theta)^2 = \mathrm{Var}(\hat{\theta}-\theta) + [E(\hat{\theta}-\theta)]^2 = \mathrm{Var}(\hat{\theta}) + [E(\hat{\theta})-\theta]^2 \text{。}$$

均方误差由估计量$\hat{\theta}$的方差$\mathrm{Var}(\hat{\theta})$、估计值与真值的偏差平方$[E(\hat{\theta})-\theta]^2$两部分组成，前者对应于有效性，后者对应于无偏性。

例 设总体X服从正态分布$N(\mu, \sigma^2)$，$X_1, X_2, \cdots, X_n$为样本。样本方差$S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$与未修正样本方差$S*^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$作为$\sigma^2$的估计量，在均方误差意义下判断其优劣。进一步问$\lambda$为何值时，$S_\lambda^2 = \lambda\sum_{i=1}^{n}(X_i-\bar{X})^2$作为$\sigma^2$的估计量均方误差最小？

解：因

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2 \sim \chi^2(n-1),$$

有$E(\chi^2) = n-1$，$\mathrm{Var}(\chi^2) = 2(n-1)$，则

$$E\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right] = (n-1)\sigma^2, \quad \mathrm{Var}\left[\sum_{i=1}^{n}(X_i-\bar{X})^2\right] = 2(n-1)\sigma^4 \text{。}$$

因$S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$，有

$$E(S^2) = \frac{1}{n-1}\cdot(n-1)\sigma^2 = \sigma^2, \quad \mathrm{Var}(S^2) = \frac{1}{(n-1)^2}\cdot 2(n-1)\sigma^4 = \frac{2}{n-1}\sigma^4,$$

则

$$\mathrm{MSE}_{\sigma^2}(S^2) = \mathrm{Var}(S^2) + [E(S^2)-\sigma^2]^2 = \frac{2}{n-1}\sigma^4 \text{。}$$

因 $S^{*2}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$，有

$$E(S^{*2})=\frac{1}{n}\cdot(n-1)\sigma^2=\frac{n-1}{n}\sigma^2,\quad \mathrm{Var}(S^{*2})=\frac{1}{n^2}\cdot 2(n-1)\sigma^4=\frac{2(n-1)}{n^2}\sigma^4,$$

则

$$\mathrm{MSE}_{\sigma^2}(S^{*2})=\mathrm{Var}(S^{*2})+[E(S^{*2})-\sigma^2]^2=\frac{2(n-1)}{n^2}\sigma^4+\frac{1}{n^2}\sigma^4=\frac{2n-1}{n^2}\sigma^4,$$

可见

$$\mathrm{MSE}_{\sigma^2}(S^{*2})=\frac{2n-1}{n^2}\sigma^4<\mathrm{MSE}_{\sigma^2}(S^2)=\frac{2}{n-1}\sigma^4,$$

故在均方误差意义下 S^{*2} 比 S^2 更好。

又因 $S_\lambda^2=\lambda\sum_{i=1}^{n}(X_i-\bar{X})^2$，有

$$E(S_\lambda^2)=(n-1)\lambda\sigma^2,\quad \mathrm{Var}(S_\lambda^2)=2(n-1)\lambda^2\sigma^4,$$

则

$$\mathrm{MSE}_{\sigma^2}(S_\lambda^2)=\mathrm{Var}(S_\lambda^2)+[E(S_\lambda^2)-\sigma^2]^2=2(n-1)\lambda^2\sigma^4+[(n-1)\lambda\sigma^2-\sigma^2]^2$$

$$=[(n-1)(n+1)\lambda^2-2(n-1)\lambda+1]\sigma^4,$$

根据二次多项式的结论可知，若

$$\lambda=\frac{2(n-1)}{2(n-1)(n+1)}=\frac{1}{n+1},$$

有 $S_\lambda^2=\frac{1}{n+1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ 的均方误差最小。

§6.4 最小方差无偏估计

先保证估计量的无偏性再考虑有效性的评价方法与考虑均方误差的评价方法是两种不同的评价体系，不能简单地认为某种评价方法更好。

这一节是先保证估计量的无偏性再考虑有效性，讨论方差最小的无偏估计。

6.4.1 最小方差无偏估计

定义 设$\hat{\theta}$是θ的一个无偏估计，如果对θ的任意一个无偏估计$\tilde{\theta}$，都有

$$\mathrm{Var}(\hat{\theta}) \le \mathrm{Var}(\tilde{\theta}),$$

则称$\hat{\theta}$是θ的一致最小方差无偏估计（Uniform Minimal Variance Unbiased Estimation，简称 UMVUE）。

定理 设θ为总体X的未知参数，$X_1, X_2, \cdots, X_n$为样本，$\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$是$\theta$的一个无偏估计，且$\mathrm{Var}(\hat{\theta}) < +\infty$，则$\hat{\theta}$是参数$\theta$的 UMVUE 的充分必要条件是对任意一个满足$E(\varphi) = 0$且$\mathrm{Var}(\varphi) < +\infty$的统计量$\varphi = \varphi(X_1, X_2, \cdots, X_n)$，都有$\mathrm{Cov}(\hat{\theta}, \varphi) = 0$。

此定理表明未知参数θ的一个无偏估计$\hat{\theta}$是θ的 UMVUE 当且仅当$\hat{\theta}$与任何期望等于 0 且方差有界的统计量不相关。

证明：充分性，设任意一个满足$E(\varphi) = 0$且$\mathrm{Var}(\varphi) < +\infty$的统计量$\varphi$都满足$\mathrm{Cov}(\hat{\theta}, \varphi) = 0$。对于$\theta$的任意一个方差有界的无偏估计$\tilde{\theta}$，记$\varphi = \tilde{\theta} - \hat{\theta}$，有$E(\varphi) = 0$且$\mathrm{Var}(\varphi) < +\infty$，可得$\mathrm{Cov}(\hat{\theta}, \varphi) = 0$。则

$$\mathrm{Var}(\tilde{\theta}) = \mathrm{Var}(\hat{\theta} + \varphi) = \mathrm{Var}(\hat{\theta}) + \mathrm{Var}(\varphi) + 2\,\mathrm{Cov}(\hat{\theta}, \varphi) = \mathrm{Var}(\hat{\theta}) + \mathrm{Var}(\varphi) \ge \mathrm{Var}(\hat{\theta}),$$

故$\hat{\theta}$是θ的 UMVUE。

必要性，设$\hat{\theta}$是θ的 UMVUE。反证法，假设存在一个统计量$\varphi = \varphi(X_1, X_2, \cdots, X_n)$满足$E(\varphi) = 0$且$\mathrm{Var}(\varphi) < +\infty$，但$\mathrm{Cov}(\hat{\theta}, \varphi) \ne 0$。对任意实数$\lambda$，记$\tilde{\theta}_\lambda = \hat{\theta} - \lambda\varphi$，有

$$E(\tilde{\theta}_\lambda) = E(\hat{\theta}) - \lambda E(\varphi) = E(\hat{\theta}) = \theta,$$

$$\mathrm{Var}(\tilde{\theta}_\lambda) = \mathrm{Var}(\hat{\theta} - \lambda\varphi) = \mathrm{Var}(\hat{\theta}) + \lambda^2\,\mathrm{Var}(\varphi) - 2\lambda\,\mathrm{Cov}(\hat{\theta}, \varphi)。$$

因$\mathrm{Cov}(\hat{\theta}, \varphi) \ne 0$且$\mathrm{Var}(\hat{\theta}) \cdot \mathrm{Var}(\varphi) \ge [\mathrm{Cov}(\hat{\theta}, \varphi)]^2$，有$\mathrm{Var}(\varphi) > 0$，取$\lambda = \dfrac{\mathrm{Cov}(\hat{\theta}, \varphi)}{\mathrm{Var}(\varphi)}$，则

$$\mathrm{Var}(\tilde{\theta}_\lambda) = \mathrm{Var}(\hat{\theta}) + \lambda^2\,\mathrm{Var}(\varphi) - 2\lambda\,\mathrm{Cov}(\hat{\theta}, \varphi) = \mathrm{Var}(\hat{\theta}) - \frac{[\mathrm{Cov}(\hat{\theta}, \varphi)]^2}{\mathrm{Var}(\varphi)} < \mathrm{Var}(\hat{\theta}),$$

即$E(\tilde{\theta}_\lambda) = \theta$，但$\mathrm{Var}(\tilde{\theta}_\lambda) < \mathrm{Var}(\hat{\theta})$，这与$\hat{\theta}$是$\theta$的 UMVUE 矛盾，故任意一个满足$E(\varphi) = 0$且$\mathrm{Var}(\varphi) < +\infty$

的统计量 $\varphi=\varphi(X_1, X_2, \cdots, X_n)$，都有 $\mathrm{Cov}(\hat{\theta}, \varphi)=0$，得证。

注： 当 $E(\varphi)=0$ 时，有 $\mathrm{Cov}(\hat{\theta}, \varphi)=E(\varphi\hat{\theta})$。故此定理结论中 $\mathrm{Cov}(\hat{\theta}, \varphi)=0$ 可变为 $E(\varphi\hat{\theta})=0$。可见要判断 $\hat{\theta}$ 是 θ 的 UMVUE，就需要判断（1）无偏性，$E(\hat{\theta})=\theta$；（2）正交性，由 $E(\varphi)=0$ 推导出 $E(\varphi\hat{\theta})=0$。

例 设总体 X 服从指数分布 $Exp\left(\frac{1}{\theta}\right)$，$X_1, X_2, \cdots, X_n$ 为样本，证明样本均值 $\bar{X}$ 是参数 θ 的 UMVUE。

证明： 因

$$E(\bar{X})=E(X)=\theta,$$

则 $\bar{X}$ 是参数 θ 的无偏估计。又因样本联合密度函数

$$p(x_1, x_2, \cdots, x_n; \theta)=\prod_{i=1}^{n}\frac{1}{\theta}e^{-\frac{x_i}{\theta}}I_{x_i>0}=\frac{1}{\theta^n}e^{-\frac{1}{\theta}\sum\limits_{i=1}^{n}x_i}I_{x_1, x_2, \cdots, x_n>0}。$$

设统计量 $\varphi=\varphi(X_1, X_2, \cdots, X_n)$ 满足

$$E(\varphi)=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\varphi\frac{1}{\theta^n}e^{-\frac{1}{\theta}\sum\limits_{i=1}^{n}x_i}dx_1\cdots dx_n=\frac{1}{\theta^n}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\varphi e^{-\frac{n\bar{x}}{\theta}}dx_1\cdots dx_n=0,$$

即

$$\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\varphi e^{-\frac{n\bar{x}}{\theta}}dx_1\cdots dx_n=0。$$

两边关于 θ 求导，得

$$\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\varphi\cdot\frac{n\bar{x}}{\theta^2}e^{-\frac{n\bar{x}}{\theta}}dx_1\cdots dx_n=\frac{n}{\theta^2}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\varphi\cdot\bar{x}e^{-\frac{n\bar{x}}{\theta}}dx_1\cdots dx_n=0$$

则

$$\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\varphi\bar{x}\cdot\frac{1}{\theta^2}e^{-\frac{n\bar{x}}{\theta}}dx_1\cdots dx_n=0$$

即 $E(\varphi\bar{X})=0$，$\mathrm{Cov}(\bar{X}, \varphi)=0$，故样本均值 $\bar{X}$ 是参数 θ 的 UMVUE。

引理 设 θ 为总体 X 的未知参数，$X_1, X_2, \cdots, X_n$ 为样本，$p=p(x_1, x_2, \cdots, x_n; \theta)$ 是样本联合质量函数或联合密度函数，$T=T(X_1, X_2, \cdots, X_n)$ 是一个统计量，观测值为 $t=T(x_1, x_2, \cdots, x_n)$。设对任意一个满足 $E(\varphi)=0$ 且 $\mathrm{Var}(\varphi)<+\infty$ 的统计量 $\varphi=\varphi(X_1, X_2, \cdots, X_n)$，都有

$$\frac{\partial}{\partial\theta}\sum_{x_1, x_2, \cdots, x_n}\varphi(x_1, x_2, \cdots, x_n)p(x_1, x_2, \cdots, x_n; \theta)=\sum_{x_1, x_2, \cdots, x_n}\varphi(x_1, x_2, \cdots, x_n)\frac{\partial p(x_1, x_2, \cdots, x_n; \theta)}{\partial\theta}$$

或

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\varphi(x_1, x_2, \cdots, x_n)p(x_1, x_2, \cdots, x_n; \theta)dx_1\cdots dx_n$$
$$=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\varphi(x_1, x_2, \cdots, x_n)\frac{\partial p(x_1, x_2, \cdots, x_n; \theta)}{\partial\theta}dx_1\cdots dx_n。$$

如果$\dfrac{\partial[c(\theta)p(x_1, x_2, \cdots, x_n;\theta)]}{\partial\theta}=[a(\theta)t+b(\theta)]p(x_1, x_2, \cdots, x_n;\theta)$，其中$a(\theta), b(\theta), c(\theta)$与样本无关，则根据$E(\varphi)=0$可得到$E(\varphi T)=0$。

证明：以连续随机变量为例。因

$$E(\varphi)=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\varphi(x_1, x_2, \cdots, x_n)p(x_1, x_2, \cdots, x_n;\theta)dx_1\cdots dx_n=0,$$

而c与样本无关，则

$$\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\varphi(x_1, x_2, \cdots, x_n)\cdot c(\theta)p(x_1, x_2, \cdots, x_n;\theta)dx_1\cdots dx_n=0,$$

两边关于θ求导，得

$$\begin{aligned}0&=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\varphi\cdot\frac{\partial(cp)}{\partial\theta}dx_1\cdots dx_n=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\varphi\cdot(at+b)pdx_1\cdots dx_n\\&=a\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\varphi tpdx_1\cdots dx_n+b\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\varphi pdx_1\cdots dx_n\\&=aE(\varphi T)+bE(\varphi)=aE(\varphi T)\end{aligned}$$

故$E(\varphi T)=0$。

如上例中总体X服从指数分布$Exp\left(\dfrac{1}{\theta}\right)$，证明样本均值$\bar{X}$是参数$\theta$的 UMVUE。

因$\bar{X}$是参数θ的无偏估计，且样本联合密度函数

$$p(x_1, x_2, \cdots, x_n;\theta)=\frac{1}{\theta^n}e^{-\frac{n\bar{x}}{\theta}}I_{x_1, x_2, \cdots, x_n>0},$$

$$\theta^n p(x_1, x_2, \cdots, x_n;\theta)=e^{-\frac{n\bar{x}}{\theta}}I_{x_1, x_2, \cdots, x_n>0},$$

则

$$\frac{\partial[\theta^n p(x_1, x_2, \cdots, x_n;\theta)]}{\partial\theta}=\frac{n\bar{x}}{\theta^2}e^{-\frac{n\bar{x}}{\theta}}I_{x_1, x_2, \cdots, x_n>0}=n\theta^{n-2}\bar{x}p(x_1, x_2, \cdots, x_n;\theta),$$

令统计量$T=\bar{X}$，$c=\theta^n$，$a=n\theta^{n-2}$，$b=0$，即$\dfrac{\partial(cp)}{\partial\theta}=(a\bar{x}+b)p$，可知根据$E(\varphi)=0$可得到$E(\varphi\bar{X})=0$。从而样本均值$\bar{X}$是参数$\theta$的 UMVUE。

例　设总体X服从泊松分布$P(\lambda)$，$X_1, X_2, \cdots, X_n$为样本，证明样本均值$\bar{X}$是参数λ的 UMVUE。

证明：因

$$E(\bar{X})=E(X)=\lambda,$$

则$\bar{X}$是参数λ的无偏估计。

又因样本联合质量函数

$$p(x_1, x_2, \cdots, x_n;\lambda)=\prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i!}e^{-\lambda}=\frac{\lambda^{\sum_{i=1}^{n}x_i}}{x_1!x_2!\cdots x_n!}e^{-n\lambda},\quad x_1, x_2, \cdots, x_n=0, 1, 2, \cdots,$$

$$e^{n\lambda}p(x_1, x_2, \cdots, x_n;\lambda)=\frac{\lambda^{n\bar{x}}}{x_1!x_2!\cdots x_n!},\quad x_1, x_2, \cdots, x_n=0, 1, 2, \cdots,$$

则

$$\frac{\partial[e^{n\lambda}p(x_1, x_2, \cdots, x_n;\lambda)]}{\partial\lambda}=\frac{n\bar{x}\lambda^{n\bar{x}-1}}{x_1!x_2!\cdots x_n!}=\frac{n\bar{x}}{\lambda}e^{n\lambda}p(x_1, x_2, \cdots, x_n;\lambda),$$

令统计量$T=\bar{X}$，$c=e^{n\lambda}$，$a=\dfrac{ne^{n\lambda}}{\lambda}$，$b=0$，即$\dfrac{\partial(cp)}{\partial\lambda}=(a\bar{x}+b)p$，可知根据$E(\varphi)=0$可得到$E(\varphi\bar{X})=0$。从而样本均值$\bar{X}$是参数$\lambda$的 UMVUE。

例　设总体X服从正态分布$N(\mu,\sigma^2)$，$X_1,X_2,\cdots,X_n$为样本，证明：

（1）样本均值$\bar{X}$是参数μ的 UMVUE。

（2）当μ已知时，$\hat{\sigma}^2=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$是$\sigma^2$的 UMVUE。

（3）当μ未知时，样本方差$S^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$是$\sigma^2$的 UMVUE。

证明：（1）因

$$E(\bar{X})=E(X)=\mu,$$

则$\bar{X}$是参数μ的无偏估计。

又因样本联合密度函数

$$p(x_1,x_2,\cdots,x_n;\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}=\frac{1}{(\sqrt{2\pi})^n\sigma^n}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2},$$

$$(\sqrt{2\pi})^n\sigma^n p(x_1,x_2,\cdots,x_n;\mu,\sigma^2)=e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2},$$

则

$$\frac{\partial[(\sqrt{2\pi})^n\sigma^n p(x_1,x_2,\cdots,x_n;\mu,\sigma^2)]}{\partial\mu}=e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}\cdot\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}2(x_i-\mu)\cdot(-1)\right]$$

$$=\frac{1}{\sigma^2}\left(\sum_{i=1}^{n}x_i-n\mu\right)e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}=\frac{1}{\sigma^2}(n\bar{x}-n\mu)(\sqrt{2\pi})^n\sigma^n p(x_1,x_2,\cdots,x_n;\mu,\sigma^2)。$$

令统计量$T=\bar{X}$，$c=(\sqrt{2\pi})^n\sigma^n$，$a=\dfrac{nc}{\sigma^2}$，$b=-\dfrac{n\mu c}{\sigma^2}$，即$\dfrac{\partial(cp)}{\partial\mu}=(a\bar{x}+b)p$，可知根据$E(\varphi)=0$可得到$E(\varphi\bar{X})=0$。从而样本均值$\bar{X}$是参数$\mu$的 UMVUE。

（2）因

$$E(\hat{\sigma}^2)=\frac{1}{n}\sum_{i=1}^{n}E[(X_i-\mu)^2]=\frac{1}{n}\cdot n\sigma^2=\sigma^2,$$

则当μ已知时，$\hat{\sigma}^2$是参数σ^2的无偏估计。又因

$$(\sqrt{2\pi})^n\sigma^n p(x_1,x_2,\cdots,x_n;\mu,\sigma^2)=e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2},$$

则

$$\frac{\partial[(\sqrt{2\pi})^n\sigma^n p(x_1,x_2,\cdots,x_n;\mu,\sigma^2)]}{\partial(\sigma^2)}=\frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2\cdot e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$=\frac{1}{2\sigma^4}\sum_{i=1}^n(x_i-\mu)^2\cdot(\sqrt{2\pi})^n\sigma^n p(x_1,x_2,\cdots,x_n;\mu,\sigma^2)\text{。}$$

令统计量 $T=\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^n(X_i-\mu)^2$ ，$c=(\sqrt{2\pi})^n\sigma^n$ ，$a=\frac{nc}{2\sigma^4}$ ，$b=0$ ，即 $\frac{\partial(cp)}{\partial(\sigma^2)}=(a\hat{\sigma}^2+b)p$ ，可知根据 $E(\varphi)=0$ 可得到 $E(\varphi\hat{\sigma}^2)=0$ 。从而当 μ 已知时，$\hat{\sigma}^2$ 是参数 σ^2 的 UMVUE。

（3）因 $E(S^2)=\sigma^2$，则 S^2 是参数 σ^2 的无偏估计。需要证明根据 $E(\varphi)=0$ 可得到 $E(\varphi S^2)=0$ 。

因

$$S^2=\frac{1}{n-1}\sum_{i=1}^n(X_i-\bar{X})^2=\frac{1}{n-1}\left(\sum_{i=1}^n X_i^2-n\bar{X}^2\right),$$

下面证明根据 $E(\varphi)=0$ 可得到 $E\left(\varphi\sum_{i=1}^n X_i^2\right)=0$ 与 $E(\varphi\bar{X}^2)=0$，从而得到 $E(\varphi S^2)=0$ 。

根据第（1）、（2）问，根据 $E(\varphi)=0$ 可得到 $E(\varphi\bar{X})=0$ 与 $E(\varphi\hat{\sigma}^2)=0$ 。

取 $\tilde{\varphi}=\varphi\bar{X}$，根据 $E(\varphi)=0$ 可得到 $E(\tilde{\varphi})=0$，再根据 $E(\tilde{\varphi})=0$ 及统计量 φ 的任意性，可得到 $E(\tilde{\varphi}\bar{X})=0$，即 $E(\varphi\bar{X}^2)=0$ 。

因

$$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^n(X_i-\mu)^2=\frac{1}{n}\sum_{i=1}^n(X_i^2-2\mu X_i+\mu^2)=\frac{1}{n}\sum_{i=1}^n X_i^2-2\mu\bar{X}+\mu^2$$

根据 $E(\varphi)=0$ 可得到 $E(\varphi\hat{\sigma}^2)=0$ ，即

$$E\left(\frac{1}{n}\varphi\sum_{i=1}^n X_i^2-2\mu\varphi\bar{X}+\varphi\mu^2\right)=\frac{1}{n}E\left(\varphi\sum_{i=1}^n X_i^2\right)-2\mu E(\varphi\bar{X})+\mu^2E(\varphi)=0\text{。}$$

因 $E(\varphi)=0$ 且 $E(\varphi\bar{X})=0$，有

$$E\left(\varphi\sum_{i=1}^n X_i^2\right)=0,$$

可得

$$E(\varphi S^2)=\frac{1}{n-1}\left[E\left(\varphi\sum_{i=1}^n X_i^2\right)-nE(\varphi\bar{X}^2)\right]=0,$$

从而当 μ 未知时，S^2 是参数 σ^2 的 UMVUE。

注：当 μ 已知时，样本联合密度函数不能关于 μ 求偏导数，就不能根据 $E(\varphi)=0$ 得到 $E(\varphi\bar{X})=0$，也就不能得到 $E(\varphi S^2)=0$，此时 S^2 就不是参数 σ^2 的 UMVUE。

6.4.2　充分性原则

定理　设θ为总体X的未知参数，$X_1, X_2, \cdots, X_n$为样本，$T = T(X_1, X_2, \cdots, X_n)$是$\theta$的充分统计量，$\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$是$\theta$的一个无偏估计，记$\tilde{\theta} = E(\hat{\theta} | T)$，则$\tilde{\theta}$也是$\theta$的一个无偏估计，且

$$\mathrm{Var}(\tilde{\theta}) \le \mathrm{Var}(\hat{\theta}),$$

其中等号成立的充分必要条件是$\tilde{\theta}$与$\hat{\theta}$几乎处处相等。

证明：因T是θ的充分统计量，条件质量函数或条件密度函数$p(x_1, x_2, \cdots, x_n | T = t)$与参数$\theta$无关，有

$$E(\hat{\theta} | T = t) = \begin{cases} \displaystyle\sum_{x_1, x_2, \cdots, x_n} \hat{\theta}(x_1, x_2, \cdots, x_n) p(x_1, x_2, \cdots, x_n | T = t), & \text{离散;} \\ \displaystyle\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} \hat{\theta}(x_1, x_2, \cdots, x_n) p(x_1, x_2, \cdots, x_n | T = t) dx_1 \cdots dx_n, & \text{连续.} \end{cases}$$

与参数θ无关，且$\tilde{\theta} = E(\hat{\theta} | T)$是统计量$T$的函数，故$\tilde{\theta} = E(\hat{\theta} | T)$是一个统计量。

根据重期望公式得

$$E(\tilde{\theta}) = E[E(\hat{\theta} | T)] = E(\hat{\theta}) = \theta$$

故$\tilde{\theta}$是θ的一个无偏估计。再根据条件方差公式得

$$\mathrm{Var}(\hat{\theta}) = \mathrm{Var}[E(\hat{\theta} | T)] + E[\mathrm{Var}(\hat{\theta} | T)] = \mathrm{Var}(\tilde{\theta}) + E[\mathrm{Var}(\hat{\theta} | T)] \ge \mathrm{Var}(\tilde{\theta}),$$

等号成立当且仅当$E[\mathrm{Var}(\hat{\theta} | T)] = 0$，而

$$\mathrm{Var}(\hat{\theta} | T) = E[(\hat{\theta} - E(\hat{\theta} | T))^2 | T] = E[(\hat{\theta} - \tilde{\theta})^2 | T],$$

$$E[\mathrm{Var}(\hat{\theta} | T)] = E\{E[(\hat{\theta} - \tilde{\theta})^2 | T]\} = E[(\hat{\theta} - \tilde{\theta})^2] = \mathrm{Var}(\hat{\theta} - \tilde{\theta}) + [E(\hat{\theta} - \tilde{\theta})]^2 = 0,$$

从而

$$\mathrm{Var}(\hat{\theta} - \tilde{\theta}) = 0 \text{ 且 } E(\hat{\theta} - \tilde{\theta}) = 0,$$

即$\hat{\theta} - \tilde{\theta}$几乎处处等于 0，故$\mathrm{Var}(\tilde{\theta}) = \mathrm{Var}(\hat{\theta})$成立的充分必要条件是$\tilde{\theta}$与$\hat{\theta}$几乎处处相等。

注：因$\tilde{\theta} = E(\hat{\theta} | T)$是$\theta$充分统计量$T$的函数，可知$\tilde{\theta} = E(\hat{\theta} | T)$也是$\theta$的充分统计量。如果$\hat{\theta}$不是$\theta$的充分统计量，则充分统计量$\tilde{\theta}$不可能与$\hat{\theta}$几乎处处相等，就有$\mathrm{Var}(\tilde{\theta}) < \mathrm{Var}(\hat{\theta})$，即$\hat{\theta}$不是$\theta$的 UMVUE。由此可知参数$\theta$的 UMVUE 必为$\theta$的充分统计量，这称之为充分性原则。

设总体X服从均匀分布$U(0, 2\mu)$，$X_1, X_2, \cdots, X_n$为样本，总体期望$E(X) = \mu$。样本均值$\bar{X}$作为$E(X) = \mu$的点估计。最大顺序统计量$X_{(n)}$是μ的充分统计量，但$\bar{X}$不是。根据此定理知$\tilde{\mu} = E(\bar{X} | X_{(n)})$是

μ 的一个无偏估计，且 $\mathrm{Var}(\tilde{\mu}) < \mathrm{Var}(\bar{X})$，故 $\bar{X}$ 不是 $E(X)$ 的 UMVUE。

6.4.3 Cramer-Rao 不等式

这一节将给出无偏估计的方差的下界，首先定义费希尔信息量。

定义 设随机变量 X 的质量函数或密度函数为 $p(x;\theta)$，参数 θ 取值范围是开区间，支撑区间与参数 θ 无关， $p(x;\theta)$ 关于 θ 可导，并且 $p(x;\theta)$ 关于 x 的积分与关于 θ 的导数可交换次序。如果

$$I(\theta) = E\left[\frac{\partial \ln p(X;\theta)}{\partial\theta}\right]^2$$

存在，则称为参数 θ 的费希尔信息量（Fisher Information）。

注： 均匀分布 $U(0,\theta)$ 的支撑区间为 $(0,\theta)$ 与参数 θ 有关，不存在费希尔信息量。

例 求泊松分布 $P(\lambda)$ 的费希尔信息量 $I(\lambda)$ 。

解： 因泊松分布 $P(\lambda)$ 的质量函数为

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

则

$$\ln p(x;\lambda) = x\ln\lambda - \lambda - \ln(x!), \quad x = 0, 1, 2, \cdots,$$
$$\frac{\partial \ln p(x;\lambda)}{\partial\lambda} = \frac{x}{\lambda} - 1 = \frac{x-\lambda}{\lambda},$$

故

$$I(\lambda) = E\left[\frac{\partial \ln p(X;\lambda)}{\partial\lambda}\right]^2 = E\left(\frac{X-\lambda}{\lambda}\right)^2 = \frac{E(X-\lambda)^2}{\lambda^2} = \frac{\mathrm{Var}(X)}{\lambda^2} = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}。$$

例 两点分布 $P\{X=1\} = p$， $P\{X=0\} = 1-p$，求其费希尔信息量 $I(p)$。

解： 因两点分布的质量函数为

$$p(x;p) = p^x(1-p)^{1-x}, \quad x = 0, 1,$$

则

$$\ln p(x;p) = x\ln p + (1-x)\ln(1-p), \quad x = 0, 1,$$

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

故

$$I(p) = E\left[\frac{X-p}{p(1-p)}\right]^2 = \frac{E(X-p)^2}{p^2(1-p)^2} = \frac{\mathrm{Var}(X)}{p^2(1-p)^2} = \frac{p(1-p)}{p^2(1-p)^2} = \frac{1}{p(1-p)}。$$

例 求指数分布 $Exp(\lambda)$ 的费希尔信息量 $I(\lambda)$ 。

解： 因指数分布 $Exp(\lambda)$ 的密度函数为

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

则

$$\ln p(x;\lambda) = \ln\lambda - \lambda x, \quad x > 0,$$
$$\frac{\partial \ln p(x;\lambda)}{\partial\lambda} = \frac{1}{\lambda} - x,$$

故

$$I(\lambda)=E\left(\frac{1}{\lambda}-X\right)^2=E\left(X-\frac{1}{\lambda}\right)^2=\mathrm{Var}(X)=\frac{1}{\lambda^2}。$$

例 求正态分布 $N(\mu,\sigma^2)$ 的费希尔信息量 $I(\mu)$ 与 $I(\sigma^2)$。

解： 因正态分布 $N(\mu,\sigma^2)$ 的密度函数为

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

则

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

$$\frac{\partial \ln p(x;\mu,\sigma^2)}{\partial\mu}=-\frac{2(x-\mu)\cdot(-1)}{2\sigma^2}=\frac{x-\mu}{\sigma^2},$$

故

$$I(\mu)=E\left(\frac{X-\mu}{\sigma^2}\right)^2=\frac{E(X-\mu)^2}{\sigma^4}=\frac{\mathrm{Var}(X)}{\sigma^4}=\frac{1}{\sigma^2}。$$

又因

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

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

则

$$I(\sigma^2)=E\left[\frac{(X-\mu)^2-\sigma^2}{2\sigma^4}\right]^2。$$

因 $X\sim N(\mu,\sigma^2)$，有 $\frac{X-\mu}{\sigma}\sim N(0,1)$，$\frac{(X-\mu)^2}{\sigma^2}\sim\chi^2(1)$，则

$$E\left[\frac{(X-\mu)^2}{\sigma^2}\right]=1,\quad \mathrm{Var}\left[\frac{(X-\mu)^2}{\sigma^2}\right]=2,$$

即

$$E[(X-\mu)^2]=\sigma^2,\quad \mathrm{Var}[(X-\mu)^2]=2\sigma^4,$$

故

$$I(\sigma^2)=E\left[\frac{(X-\mu)^2-\sigma^2}{2\sigma^4}\right]^2=\frac{E[(X-\mu)^2-E(X-\mu)^2]^2}{4\sigma^8}=\frac{\mathrm{Var}[(X-\mu)^2]}{4\sigma^8}=\frac{2\sigma^4}{4\sigma^8}=\frac{1}{2\sigma^4}。$$

定理 （Cramer-Rao 不等式）设总体 X 的质量函数或密度函数为 $p(x;\theta)$，满足费希尔信息量存在的条件，$I(\theta)$ 是其费希尔信息量。$X_1,X_2,\cdots,X_n$ 为样本，$T=T(X_1,X_2,\cdots,X_n)$ 是 $g(\theta)$ 的任意一个无偏估计，$g'(\theta)$ 存在，且 $g(\theta)=E(T)$ 关于 θ 求导与关于 $x_1,x_2,\cdots,x_n$ 积分可交换顺序，则

$$\mathrm{Var}(T)\ge\frac{[g'(\theta)]^2}{nI(\theta)}。$$

证明：以连续型情形为例证明。因

$$\int_{-\infty}^{+\infty}p(x_i;\theta)dx_i=1,\quad i=1,2,\cdots,n,$$

关于θ求偏导数，并注意到

$$\frac{\partial\ln p(x_i;\theta)}{\partial\theta}=\frac{1}{p(x_i;\theta)}\frac{\partial p(x_i;\theta)}{\partial\theta},$$

则

$$0=\frac{\partial}{\partial\theta}\int_{-\infty}^{+\infty}p(x_i;\theta)dx_i=\int_{-\infty}^{+\infty}\frac{\partial p(x_i;\theta)}{\partial\theta}dx_i=\int_{-\infty}^{+\infty}\frac{\partial\ln p(x_i;\theta)}{\partial\theta}p(x_i;\theta)dx_i=E\left[\frac{\partial\ln p(X_i;\theta)}{\partial\theta}\right]。$$

记

$$Z=\frac{\partial}{\partial\theta}\ln p(X_1,X_2,\cdots,X_n;\theta)=\frac{\partial}{\partial\theta}\ln\prod_{i=1}^{n}p(X_i;\theta)=\sum_{i=1}^{n}\frac{\partial\ln p(X_i;\theta)}{\partial\theta},$$

则

$$E(Z)=\sum_{i=1}^{n}E\left[\frac{\partial\ln p(X_i;\theta)}{\partial\theta}\right]=0,$$

$$\mathrm{Var}(Z)=\sum_{i=1}^{n}\mathrm{Var}\left[\frac{\partial\ln p(X_i;\theta)}{\partial\theta}\right]=\sum_{i=1}^{n}E\left[\frac{\partial\ln p(X_i;\theta)}{\partial\theta}\right]^2=nI(\theta)。$$

因$T=T(X_1,X_2,\cdots,X_n)$是$g(\theta)$的任意一个无偏估计，有

$$g(\theta)=E(T)=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}T(x_1,x_2,\cdots,x_n)p(x_1,x_2,\cdots,x_n;\theta)dx_1\cdots dx_n$$

则

$$\begin{aligned}g'(\theta)&=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}T(x_1,x_2,\cdots,x_n)\frac{\partial}{\partial\theta}p(x_1,x_2,\cdots,x_n;\theta)dx_1\cdots dx_n\\&=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}T(x_1,x_2,\cdots,x_n)\frac{\partial}{\partial\theta}\ln p(x_1,x_2,\cdots,x_n;\theta)\cdot p(x_1,x_2,\cdots,x_n;\theta)dx_1\cdots dx_n\\&=E(TZ)。\end{aligned}$$

因$E(Z)=0$，则

$$[g'(\theta)]^2=[E(TZ)]^2=[\mathrm{Cov}(T,Z)]^2\le\mathrm{Var}(T)\mathrm{Var}(Z)=\mathrm{Var}(T)\cdot nI(\theta),$$

故

$$\mathrm{Var}(T)\ge\frac{[g'(\theta)]^2}{nI(\theta)}。$$

注：（1）$\dfrac{[g'(\theta)]^2}{nI(\theta)}$称为$g(\theta)$的 C-R 下界。如果$g(\theta)$的一个无偏估计$T$，满足$\mathrm{Var}(T)=\dfrac{[g'(\theta)]^2}{nI(\theta)}$，则称$T$是$g(\theta)$的有效估计。显然有效估计必为 UMVUE，但是有效估计并不一定总是存在。

（2）$\mathrm{Var}(T)=\dfrac{[g'(\theta)]^2}{nI(\theta)}$ 等式成立的充分必要条件是 $[\mathrm{Cov}(T,Z)]^2=\mathrm{Var}(T)\mathrm{Var}(Z)$，$|\mathrm{Corr}(T,Z)|=1$，即 T 与 Z 几乎必然有线性关系，存在数 a,b，使得 $Z=aT+b$，即

$$Z=\frac{\partial}{\partial\theta}\ln p(X_1,X_2,\cdots,X_n;\theta)=\frac{1}{p(X_1,X_2,\cdots,X_n;\theta)}\frac{\partial}{\partial\theta}p(X_1,X_2,\cdots,X_n;\theta)=aT+b\text{，a.e.，}$$

$$\frac{\partial}{\partial\theta}p(X_1,X_2,\cdots,X_n;\theta)=(aT+b)p(X_1,X_2,\cdots,X_n;\theta)\text{，a.e.，}$$

这与前面判断 UMVUE 的结果类似。

（3）令 $g(\theta)=\theta$，可得对于 θ 的任意一个无偏估计 $\hat{\theta}$，有 $\mathrm{Var}(\hat{\theta})\ge\dfrac{1}{nI(\theta)}$，即 $\dfrac{1}{nI(\theta)}$ 为 θ 的 C-R 下界。

若 θ 的一个无偏估计 $\hat{\theta}$ 满足 $\mathrm{Var}(\hat{\theta})=\dfrac{1}{nI(\theta)}$，则 $\hat{\theta}$ 是 θ 的有效估计。

（4）因

$$I[g(\theta)]=E\left[\frac{\partial\ln p(X;\theta)}{\partial g(\theta)}\right]^2=E\left[\frac{\partial\ln p(X;\theta)}{\partial\theta}\Big/\frac{dg(\theta)}{d\theta}\right]^2=\frac{E\left[\dfrac{\partial\ln p(X;\theta)}{\partial\theta}\right]^2}{[g'(\theta)]^2}=\frac{I(\theta)}{[g'(\theta)]^2}\text{，}$$

可见 $g(\theta)$ 的 C-R 下界 $\dfrac{[g'(\theta)]^2}{nI(\theta)}$ 也可写成 $\dfrac{1}{nI[g(\theta)]}$。

例　设总体 X 服从泊松分布 $P(\lambda)$，$X_1,X_2,\cdots,X_n$ 为样本，证明样本均值 $\bar{X}$ 是参数 λ 的有效估计。

证明：因

$$E(\bar{X})=E(X)=\lambda\text{，}\quad \mathrm{Var}(\bar{X})=\frac{\mathrm{Var}(X)}{n}=\frac{\lambda}{n}\text{，}$$

则 $\bar{X}$ 是参数 λ 的无偏估计。又因 λ 的费希尔信息量 $I(\lambda)=\dfrac{1}{\lambda}$，C-R 下界为 $\dfrac{1}{nI(\lambda)}=\dfrac{\lambda}{n}$，故

$$\mathrm{Var}(\bar{X})=\frac{\lambda}{n}=\frac{1}{nI(\lambda)}\text{，}$$

即样本均值 $\bar{X}$ 是参数 λ 的有效估计。

例　设总体 X 服从指数分布 $Exp(\lambda)$，$X_1,X_2,\cdots,X_n$ 为样本，证明样本均值 $\bar{X}$ 是 $g(\lambda)=\dfrac{1}{\lambda}$ 的有效估计。

证明：因

$$E(\bar{X})=E(X)=\frac{1}{\lambda}\text{，}\quad \mathrm{Var}(\bar{X})=\frac{\mathrm{Var}(X)}{n}=\frac{1}{n\lambda^2}\text{，}$$

则 $\bar{X}$ 是 $g(\lambda)=\dfrac{1}{\lambda}$ 的无偏估计。又因 λ 的费希尔信息量 $I(\lambda)=\dfrac{1}{\lambda^2}$，$g(\lambda)=\dfrac{1}{\lambda}$ 的 C-R 下界为 $\dfrac{[g'(\lambda)]^2}{nI(\lambda)}=\dfrac{1}{n\lambda^2}$，

故

$$\mathrm{Var}(\bar{X})=\frac{1}{n\lambda^2}=\frac{[g'(\lambda)]^2}{nI(\lambda)},$$

即样本均值 $\bar{X}$ 是 $g(\lambda)=\dfrac{1}{\lambda}$ 的有效估计。

例 设总体 X 服从正态分布 $N(\mu,\sigma^2)$，$X_1,X_2,\cdots,X_n$ 为样本，证明样本均值 $\bar{X}$ 是参数 μ 的有效估计，但样本方差 S^2 不是参数 σ^2 的有效估计。

证明：因

$$E(\bar{X})=E(X)=\mu,\quad \mathrm{Var}(\bar{X})=\frac{\mathrm{Var}(X)}{n}=\frac{\sigma^2}{n},$$

则 $\bar{X}$ 是参数 μ 的无偏估计。又因 μ 的费希尔信息量 $I(\mu)=\dfrac{1}{\sigma^2}$，C-R 下界为 $\dfrac{1}{nI(\mu)}=\dfrac{\sigma^2}{n}$，故

$$\mathrm{Var}(\bar{X})=\frac{\sigma^2}{n}=\frac{1}{nI(\mu)},$$

即样本均值 $\bar{X}$ 是参数 μ 的有效估计。

因 $\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$，有

$$E\left[\frac{(n-1)S^2}{\sigma^2}\right]=n-1,\quad \mathrm{Var}\left[\frac{(n-1)S^2}{\sigma^2}\right]=2(n-1),$$

$$E(S^2)=\sigma^2,\quad \mathrm{Var}(S^2)=\frac{2\sigma^4}{n-1},$$

则 S^2 是参数 σ^2 的无偏估计。又因 σ^2 的费希尔信息量 $I(\sigma^2)=\dfrac{1}{2\sigma^4}$，C-R 下界为 $\dfrac{1}{nI(\sigma^2)}=\dfrac{2\sigma^4}{n}$，故

$$\mathrm{Var}(S^2)=\frac{2\sigma^4}{n-1}>\frac{2\sigma^4}{n}=\frac{1}{nI(\sigma^2)},$$

即样本方差 S^2 不是参数 σ^2 的有效估计。

注：正态总体 $X\sim N(\mu,\sigma^2)$ 中，样本方差 S^2 是参数 σ^2 的 UMVUE，但不是有效估计，可见 σ^2 的有效估计不存在。

定义 设对于随机变量 $\hat{\theta}_n$，存在常数 θ 及趋于 0 的常数序列 $\sigma_n(\theta)$，使得 $\dfrac{\hat{\theta}_n-\theta}{\sigma_n(\theta)}$ 依分布收敛于标准正态分布，则称 $\hat{\theta}_n$ 服从渐近正态分布 $N(\theta,\sigma_n^2(\theta))$，记为 $\hat{\theta}_n\sim AN(\theta,\sigma_n^2(\theta))$，并称 $\sigma_n^2(\theta)$ 为 $\hat{\theta}_n$ 的渐近方差。

定理 设总体 X 的质量函数或密度函数为 $p(x;\theta)$，参数 θ 取值范围为非退化区间。假定 $\ln p(x;\theta)$ 对任意的 x 关于 θ 三阶可导，$\dfrac{\partial p(x;\theta)}{\partial\theta}$，$\dfrac{\partial^2 p(x;\theta)}{\partial\theta^2}$ 关于 x 在 $(-\infty,+\infty)$ 上求和或积分以及 $\dfrac{\partial^3\ln p(X;\theta)}{\partial\theta^3}$ 的期望对任意的 θ 都一致有界，且满足费希尔信息量存在的条件，$I(\theta)$ 是费希尔信息量。$X_1,X_2,\cdots,X_n$ 为样

本，则存在参数θ的最大似然估计$\hat{\theta}_n=\hat{\theta}_n(X_1, X_2, \cdots, X_n)$具有相合性和渐近正态性，$\hat{\theta}_n \sim AN\left(\theta, \dfrac{1}{nI(\theta)}\right)$，

即$\hat{\theta}_n$的渐近方差为参数θ的 C-R 下界。

§6.5　贝叶斯估计

传统上，统计学经典学派认为总体的未知参数θ是一个未知的确定数，通过样本建立估计量去估计θ的真实值；而贝叶斯学派认为θ是一个随机变量，其分布称为先验分布，通过样本对其分布作出重新判断，得到θ的后验分布，再根据后验分布作出对θ的估计。

6.5.1　　统计推断的基础

统计学经典学派进行统计推断需利用总体信息、样本信息，即总体分布、样本观测值的信息；而贝叶斯学派进行统计推断时，除了利用以上两种信息之外，还将利用先验信息，即参数的先验分布。

可见，贝叶斯学派进行统计推断利用的信息更加丰富。

6.5.2　　贝叶斯公式的密度函数形式

条件概率的贝叶斯公式：设$B_1, B_2, \cdots, B_n$为样本空间Ω的一个分割，A为任一随机事件，$P(A)>0$，$P(B_i)>0,\quad i=1,2,\cdots,n$，则有

$$P(B_k\mid A)=\frac{P(AB_k)}{P(A)}=\frac{P(B_k)P(A\mid B_k)}{\sum_{i=1}^{n}P(B_i)P(A\mid B_i)}。$$

实际背景：$P(B_k)$为先验概率，反映检验（事件A）之前，根据经验判断B_k发生的概率；而$P(B_k\mid A)$为后验概率，根据检验的结果对B_k发生的概率作出重新判断。

设二维连续随机变量(X,Y)的联合密度函数为$p(x,y)$，有条件密度的公式

$$p_Y(y\mid X=x)=\frac{p(x,y)}{p_X(x)}=\frac{p(x,y)}{\int_{-\infty}^{+\infty}p(x,y)dy}=\frac{p_Y(y)p_X(x\mid Y=y)}{\int_{-\infty}^{+\infty}p_Y(y)p_X(x\mid Y=y)dy},$$

这称为贝叶斯公式的密度函数形式，其中Y的边际密度$p_Y(y)$为Y的先验分布，条件密度$p_Y(y\mid X=x)$为Y的后验分布。

统计学贝叶斯的学派认为总体X的分布中所含未知参数θ是一个随机变量，参数空间为Θ。

参数θ不同取值相当于贝叶斯公式密度函数形式中的随机变量Y，样本$X_1, X_2, \cdots, X_n$相当于随机变量X。根据θ的先验分布，推知其后验分布，步骤：

（1）给出未知参数θ的先验分布$\pi(\theta)$。

这对应于贝叶斯公式密度函数形式中随机变量Y的先验分布$p_Y(y)$。

（2）根据总体条件分布$p(x\mid\theta)$得到样本条件分布$p(x_1, x_2, \cdots, x_n\mid\theta)=\prod_{i=1}^{n}p(x_i\mid\theta)$。传统上的总体分布$p(x;\theta)$和样本联合分布$p(x_1, x_2, \cdots, x_n;\theta)$在贝叶斯学派看来是参数$\theta$取给定值时的总体条件分布$p(x\mid\theta)$和样本条件分布$p(x_1, x_2, \cdots, x_n\mid\theta)$。

这里样本条件分布$p(x_1, x_2, \cdots, x_n\mid\theta)$对应于贝叶斯公式密度函数形式中的条件分布$p_X(x\mid Y=y)$。

（3）样本与参数联合分布 $h(x_1, x_2, \cdots, x_n, \theta) = \pi(\theta) p(x_1, x_2, \cdots, x_n \mid \theta)$。

这对应于贝叶斯公式密度函数形式中的分子 $p(x, y) = p_Y(y) p_X(x \mid Y = y)$。

（4）样本边际分布 $m(x_1, x_2, \cdots, x_n) = \int_\Theta h(x_1, x_2, \cdots, x_n, \theta) d\theta = \int_\Theta \pi(\theta) p(x_1, x_2, \cdots, x_n \mid \theta) d\theta$。

这对应于贝叶斯公式密度函数形式中的分母 $p_X(x) = \int_{-\infty}^{+\infty} p(x, y) dy = \int_{-\infty}^{+\infty} p_Y(y) p_X(x \mid Y = y) dy$。

（5）得到参数 θ 的后验分布 $\pi(\theta \mid x_1, x_2, \cdots, x_n) = \dfrac{h(x_1, x_2, \cdots, x_n, \theta)}{m(x_1, x_2, \cdots, x_n)} = \dfrac{\pi(\theta) p(x_1, x_2, \cdots, x_n \mid \theta)}{\int_\Theta \pi(\theta) p(x_1, x_2, \cdots, x_n \mid \theta) d\theta}$。

这对应于贝叶斯公式密度函数形式 $p_Y(y \mid X = x) = \dfrac{p(x, y)}{p_X(x)} = \dfrac{p_Y(y) p_X(x \mid Y = y)}{\int_{-\infty}^{+\infty} p_Y(y) p_X(x \mid Y = y) dy}$。

6.5.3　　贝叶斯估计

通常以参数 θ 后验分布 $\pi(\theta \mid x_1, x_2, \cdots, x_n)$ 的数学期望 $E(\theta \mid X_1, X_2, \cdots, X_n)$ 作为 θ 的点估计，称为后验期望估计，又称为贝叶斯估计，记为 $\hat{\theta}_B$。

例　设事件 A 在一次试验中发生的概率为 p，进行 n 次独立重复试验，X 表示事件 A 的发生次数，求概率 p 的贝叶斯估计。

解：试验之前对事件 A 发生的概率并无了解，以均匀分布 $U(0,1)$ 作为概率 p 的先验分布，

$$\pi(p) = I_{0<p<1}\text{。}$$

设总体为每一次试验中事件 A 的发生次数，总体服从两点分布 $b(1, p)$，总体条件分布

$$p(x \mid p) = p^x (1-p)^{1-x}, \quad x = 0, 1,$$

样本条件分布

$$p(x_1, x_2, \cdots, x_n \mid p) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{\sum\limits_{i=1}^n x_i} (1-p)^{n - \sum\limits_{i=1}^n x_i}, \quad x_1, x_2, \cdots, x_n = 0, 1,$$

事件 A 的发生次数 $x = \sum\limits_{i=1}^n x_i$，有

$$p(x_1, x_2, \cdots, x_n \mid p) = p^x (1-p)^{n-x}, \quad x_1, x_2, \cdots, x_n = 0, 1\text{。}$$

分子为样本与参数联合分布

$$h(x_1, x_2, \cdots, x_n, p) = \pi(p) p(x_1, x_2, \cdots, x_n \mid p) = p^x (1-p)^{n-x} I_{0<p<1}\text{。}$$

分母为样本边际分布

$$m(x_1, x_2, \cdots, x_n) = \int_0^1 p^x (1-p)^{n-x} dp = \beta(x+1, n-x+1) = \frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)}\text{。}$$

参数 p 的后验分布

$$\pi(p \mid x_1, x_2, \cdots, x_n) = \frac{h(x_1, x_2, \cdots, x_n, p)}{m(x_1, x_2, \cdots, x_n)} = \frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)} p^x (1-p)^{n-x} I_{0<p<1},$$

这是贝塔分布 $Be(x+1, n-x+1)$，期望为

$$E(p\,|\,x_1, x_2, \cdots, x_n) = \int_0^1 p\cdot\frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}p^x(1-p)^{n-x}dp$$

$$=\frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}\int_0^1 p^{x+1}(1-p)^{n-x}dp$$

$$=\frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}\cdot\beta(x+2, n-x+1)$$

$$=\frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}\cdot\frac{\Gamma(x+2)\Gamma(n-x+1)}{\Gamma(n+3)}=\frac{x+1}{n+2},$$

故 p 的贝叶斯估计

$$\hat{p}_B = E(p\,|\,X_1, X_2, \cdots, X_n) = \frac{X+1}{n+2}。$$

注：（1）此题也可将每 n 次试验的结果作为个体，则总体服从二项分布，总体条件分布为

$$p(x\,|\,p) = C_n^x p^x(1-p)^{n-x},\quad x = 0, 1, \cdots, n,$$

这种情况下 n 次独立重复试验的结果看作容量为 1 的样本，样本条件分布也为 $p(x\,|\,p)$。

（2）此题条件下，传统上概率 p 的点估计为频率 $\hat{p}=\frac{X}{n}$，不同于贝叶斯估计 $\hat{p}_B=\frac{X+1}{n+2}$。在某些极端情形下贝叶斯估计显得更加“合理”。如 3 次试验发生 3 次，频率 $\hat{p}=1$，贝叶斯估计 $\hat{p}_B=0.8$；而 8 次试验发生 8 次，频率 $\hat{p}=1$，贝叶斯估计 $\hat{p}_B=0.9$。后者显然比前者更加可靠，但从频率看二者相同，贝叶斯估计就能体现出不同。

例 设总体 X 服从指数分布 $Exp(\lambda)$，参数 λ 的先验分布也是指数分布 $Exp(\theta)$，θ 已知，$X_1, X_2, \cdots, X_n$ 为样本，求 λ 的贝叶斯估计 $\hat{\lambda}_B$。

解：参数 λ 的先验分布为

$$\pi(\lambda) = \theta e^{-\theta\lambda}I_{\lambda>0}。$$

总体条件分布

$$p(x\,|\,\lambda) = \lambda e^{-\lambda x}I_{x>0},$$

样本条件分布

$$p(x_1, x_2, \cdots, x_n\,|\,\lambda) = \prod_{i=1}^n \lambda e^{-\lambda x_i}I_{x_i>0} = \lambda^n e^{-\lambda\sum_{i=1}^n x_i}I_{x_1, x_2, \cdots, x_n>0},$$

分子为样本与参数联合分布

$$h(x_1, x_2, \cdots, x_n, \lambda) = \pi(\lambda)p(x_1, x_2, \cdots, x_n\,|\,\lambda) = \theta\lambda^n e^{-\lambda\left(\theta+\sum_{i=1}^n x_i\right)}I_{x_1, x_2, \cdots, x_n, \lambda>0},$$

分母为样本边际分布

$$m(x_1, x_2, \cdots, x_n)=\int_0^{+\infty}\theta\lambda^n e^{-\lambda\left(\theta+\sum_{i=1}^n x_i\right)} I_{x_1, x_2, \cdots, x_n}\,d\lambda=\frac{\Gamma(n+1)}{\left(\theta+\sum_{i=1}^n x_i\right)^{n+1}}\theta I_{x_1, x_2, \cdots, x_n}\text{。}$$

参数λ的后验分布

$$\pi(\lambda\,|\,x_1, x_2, \cdots, x_n)=\frac{h(x_1, x_2, \cdots, x_n, \lambda)}{m(x_1, x_2, \cdots, x_n)}=\frac{\left(\theta+\sum_{i=1}^n x_i\right)^{n+1}}{\Gamma(n+1)}\lambda^n e^{-\lambda\left(\theta+\sum_{i=1}^n x_i\right)} I_{\lambda>0}\,,$$

这是伽玛分布$Ga\left(n+1, \theta+\sum_{i=1}^n x_i\right)$，期望为

$$E(\lambda\,|\,x_1, x_2, \cdots, x_n)=\int_0^{+\infty}\lambda\cdot\frac{\left(\theta+\sum_{i=1}^n x_i\right)^{n+1}}{\Gamma(n+1)}\lambda^n e^{-\lambda\left(\theta+\sum_{i=1}^n x_i\right)}d\lambda=\frac{\left(\theta+\sum_{i=1}^n x_i\right)^{n+1}}{\Gamma(n+1)}\int_0^{+\infty}\lambda^{n+1}e^{-\lambda\left(\theta+\sum_{i=1}^n x_i\right)}d\lambda$$

$$=\frac{\left(\theta+\sum_{i=1}^n x_i\right)^{n+1}}{\Gamma(n+1)}\cdot\frac{\Gamma(n+2)}{\left(\theta+\sum_{i=1}^n x_i\right)^{n+2}}=\frac{n+1}{\theta+\sum_{i=1}^n x_i}\,,$$

故λ的贝叶斯估计

$$\hat{\lambda}_B=E(\lambda\,|\,X_1, X_2, \cdots, X_n)=\frac{n+1}{\theta+\sum_{i=1}^n X_i}\text{。}$$

注：此题条件下，传统上参数λ的点估计为$\hat{\lambda}=\dfrac{1}{\overline{X}}=\dfrac{n}{\sum_{i=1}^n X_i}$，不同于贝叶斯估计$\hat{\lambda}_B=\dfrac{n+1}{\theta+\sum_{i=1}^n X_i}$。

6.5.4　共轭先验分布

定义　设θ是总体参数，若根据任意的样本观测值得到的后验分布$\pi(\theta\,|\,x_1, x_2, \cdots, x_n)$与先验分布$\pi(\theta)$属于同一个分布族，则称该分布族是$\theta$的共轭先验分布族。

如前面例子中估计事件A在一次试验中发生的概率p，进行n次独立重复试验，X表示事件A的发生次数。如果概率p的先验分布为均匀分布$U(0,1)$也就是贝塔分布$Be(1,1)$，可得p的后验分布为贝塔分布$Be(x+1, n-x+1)$。进一步可以验证如果概率p的先验分布为贝塔分布$Be(a,b)$，可得p的后验分布为贝塔分布$Be(x+a, n-x+b)$，故贝塔分布是概率p的共轭先验分布族。

又如前面例子中估计指数分布$Exp(\lambda)$中的未知参数λ，$X_1, X_2, \cdots, X_n$为样本。如果参数λ的先验分布是指数分布$Exp(\theta)$也就是伽玛分布$Ga(1,\theta)$，可得λ的后验分布为伽玛分布$Ga\left(n+1, \theta+\sum_{i=1}^n x_i\right)$。进一步可以验证如果$\lambda$的先验分布是伽玛分布$Ga(\alpha,\theta)$，可得$\lambda$的后验分布为伽玛分布$Ga\left(n+\alpha, \theta+\sum_{i=1}^n x_i\right)$，故伽玛分布是指数分布中参数$\lambda$的共轭先验分布族。

§6.6 区间估计

对参数θ作出点估计$\hat{\theta}$，可以判断一般情况下θ应在$\hat{\theta}$附近。进一步问θ究竟应离$\hat{\theta}$多近才合理？这就需要区间估计，也就是估计θ所在的一个“合理”范围。

6.6.1 区间估计的概念

一般是在$\hat{\theta}$左右各给出一个估计量$\hat{\theta}_L,\hat{\theta}_U$，估计$\theta\in[\hat{\theta}_L,\hat{\theta}_U]$。但并不能完全保证$\theta$在区间$[\hat{\theta}_L,\hat{\theta}_U]$上。通常要求$\theta$以很大的概率位于区间$[\hat{\theta}_L,\hat{\theta}_U]$上，即概率$P\{\hat{\theta}_L\le\theta\le\hat{\theta}_U\}$接近于 1。

定义 设θ是总体X的未知参数，对给定的数$\alpha\in(0,1)$，若统计量$\hat{\theta}_L,\hat{\theta}_U$满足概率

$$P\{\hat{\theta}_L\le\theta\le\hat{\theta}_U\}\ge 1-\alpha,$$

则称$[\hat{\theta}_L,\hat{\theta}_U]$是$\theta$的$1-\alpha$置信区间，相应称$1-\alpha$为置信度，$\hat{\theta}_L,\hat{\theta}_U$分别为置信下限和置信上限。

若满足$P\{\hat{\theta}_L\le\theta\le\hat{\theta}_U\}=1-\alpha$，则称$[\hat{\theta}_L,\hat{\theta}_U]$是$\theta$的$1-\alpha$同等置信区间，也称为$1-\alpha$置信区间。

通常当总体分布为连续型时，总是取同等置信区间，只有当总体分布为离散型时，一般不能使得θ位于置信区间内的概率恰好等于$1-\alpha$，才考虑大于$1-\alpha$的情形。

此外，有时只需要考虑单侧的置信区间。

定义 设θ是总体X的未知参数，给定的数$\alpha\in(0,1)$。若统计量$\hat{\theta}_L$满足$P\{\hat{\theta}_L\le\theta\}\ge1-\alpha$，则称$[\hat{\theta}_L,+\infty)$是$\theta$的$1-\alpha$单侧置信区间，$\hat{\theta}_L$为单侧置信下限；若统计量$\hat{\theta}_U$满足$P\{\theta\le\hat{\theta}_U\}=1-\alpha$，则称$(-\infty,\hat{\theta}_U]$是$\theta$的$1-\alpha$单侧置信区间，$\hat{\theta}_U$为单侧置信上限。

类似地，若θ位于置信区间内的概率等于$1-\alpha$，则相应称为同等置信区间。

本节主要是分析正态总体参数的置信区间，以下都是考虑同等置信区间。

6.6.2 枢轴量法

关于样本和未知参数θ的函数$G=G(X_1,X_2,\cdots,X_n,\theta)$，如果$G$的分布不依赖于未知参数$\theta$，则称之为参数$\theta$的枢轴量。

注：枢轴量与统计量不同，统计量是形式上不含未知参数，但其分布往往依赖于未知参数；而枢轴量却是形式上含有未知参数，但其分布不依赖未知参数。

如总体X服从正态分布$N(\mu,\sigma^2)$，参数μ未知，σ^2已知。$\bar{X}$是一个统计量，形式上不含未知参数μ，但其分布为$N\left(\mu,\dfrac{\sigma^2}{n}\right)$，依赖于$\mu$；而$\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$是一个枢轴量，形式上含有未知参数$\mu$，但其分布为$N(0,1)$，不依赖于未知参数。

计算未知参数置信区间的方法通常采用枢轴量法，步骤：

（1）给出未知参数θ的点估计$\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$，其分布依赖于$\theta$。

（2）根据 $\hat{\theta}$ 构造一个关于未知参数 θ 的枢轴量 $G = G(X_1, X_2, \cdots, X_n, \theta)$。

（3）根据置信度 $1-\alpha$ 选取常数 c, d，使得概率 $P\{c \le G \le d\} \ge 1-\alpha$。

（4）求解不等式 $c \le G \le d$，解得 $\hat{\theta}_L \le \theta \le \hat{\theta}_U$，即得到 $1-\alpha$ 置信区间 $[\hat{\theta}_L, \hat{\theta}_U]$。

通常选取常数 c, d，使得 $P\{G < c\} \le \dfrac{\alpha}{2}$ 且 $P\{G > d\} \le \dfrac{\alpha}{2}$（连续型总体是 $P\{G < c\} = P\{G > d\} = \dfrac{\alpha}{2}$，离散型总体是使得不等式成立的最大的 c 和最小的 d），由此得到的置信区间，称为等尾置信区间；此外也可以选取常数 c, d，使得置信区间 $[\hat{\theta}_L, \hat{\theta}_U]$ 的平均长度 $E(\hat{\theta}_U - \hat{\theta}_L)$ 最短，称为最短置信区间。本书都是采用等尾置信区间。

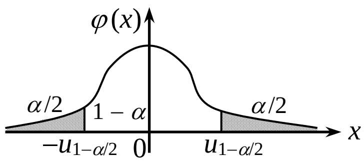
6.6.3　单个正态总体参数的置信区间

正态总体 $N(\mu, \sigma^2)$，$X_1, X_2, \cdots, X_n$ 为样本，估计均值 μ 或方差 σ^2。

一．已知方差 σ^2，估计均值 μ

步骤：

（1）用样本均值 $\bar{X}$ 作为 μ 的点估计。



（2）建立枢轴量 $U = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$。

（3）置信度 $1-\alpha$，即 $P\left\{-u_{1-\alpha/2} \le \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \le u_{1-\alpha/2}\right\} = 1-\alpha$。

（4）解得 $\bar{X} - u_{1-\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + u_{1-\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}}$。

可得 μ 的 $1-\alpha$ 置信区间 $\left[\bar{X} - u_{1-\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}}, \bar{X} + u_{1-\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}}\right]$，简记为 $\left[\bar{X} \pm u_{1-\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}}\right]$。

例　已知某种袋装食品的重量（克）服从 $N(\mu, \sigma^2)$，已知 $\sigma^2 = 16$。现随机抽取 10 袋，重量为 495, 503, 494, 502, 497, 499, 506, 500, 492, 501，求期望 μ 的 95%与 99%置信区间。

解：单个正态总体，已知 σ^2，估计 μ，用样本均值 $\bar{X}$ 作为 μ 的点估计，枢轴量 $U = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$，

则 μ 的 $1-\alpha$ 置信区间 $\mu \in \left[\bar{X} \pm u_{1-\alpha/2} \cdot \dfrac{\sigma}{\sqrt{n}}\right]$。

因 $\bar{x} = 498.9$，$\sigma = 4$，$n = 10$，置信度 $1-\alpha = 0.95$ 时，$u_{1-\alpha/2} = u_{0.975} = 1.96$，故 μ 的 95%置信区间为

$$\left[498.9 \pm 1.96 \times \frac{4}{\sqrt{10}}\right] = [496.42, 501.38]。$$

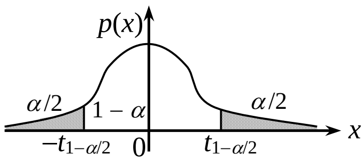
置信度 $1-\alpha = 0.99$ 时，$u_{1-\alpha/2} = u_{0.995} = 2.58$，故 μ 的 99%置信区间为

$$\left[498.9 \pm 2.58 \times \frac{4}{\sqrt{10}}\right] = [495.64, 502.16]\text{。}$$

二．未知方差σ^2，估计均值μ

已知σ^2，估计μ，用$U = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$；未知$\sigma^2$，估计$\mu$，用样本标准差$S$替换总体标准差$\sigma$，得$T = \dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$。

步骤：



（1）用样本均值$\bar{X}$作为μ的点估计。

（2）建立枢轴量$T = \dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$。

（3）置信度$1-\alpha$，即$P\left\{-t_{1-\alpha/2}(n-1) \le \dfrac{\bar{X}-\mu}{S/\sqrt{n}} \le t_{1-\alpha/2}(n-1)\right\} = 1-\alpha$。

（4）解得$\bar{X} - t_{1-\alpha/2}(n-1)\cdot\dfrac{S}{\sqrt{n}} \le \mu \le \bar{X} + t_{1-\alpha/2}(n-1)\cdot\dfrac{S}{\sqrt{n}}$。

可得μ的$1-\alpha$置信区间$\left[\bar{X} - t_{1-\alpha/2}(n-1)\cdot\dfrac{S}{\sqrt{n}}, \bar{X} + t_{1-\alpha/2}(n-1)\cdot\dfrac{S}{\sqrt{n}}\right]$，简记为$\left[\bar{X} \pm t_{1-\alpha/2}(n-1)\cdot\dfrac{S}{\sqrt{n}}\right]$。

例　前例中，若改为未知σ^2，同样的样本，求μ的 95%，99%置信区间。

解：单个正态总体，未知σ^2，估计μ，用样本均值$\bar{X}$作为μ的点估计，枢轴量$T = \dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$，则$\mu$的$1-\alpha$置信区间为$\left[\bar{X} \pm t_{1-\alpha/2}(n-1)\cdot\dfrac{S}{\sqrt{n}}\right]$。

因$\bar{x} = 498.9$，$s^2 = 4.38^2$，$n = 10$，置信度$1-\alpha = 0.95$时，$t_{1-\alpha/2}(n-1) = t_{0.975}(9) = 2.2622$，故$\mu$的 95%置信区间为

$$\left[498.9 \pm 2.2622 \times \frac{4.38}{\sqrt{10}}\right] = [495.77, 502.03]\text{。}$$

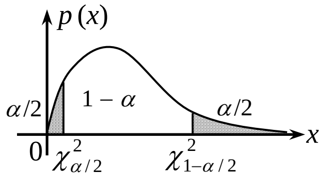
置信度$1-\alpha = 0.99$时，$t_{1-\alpha/2}(n-1) = t_{0.995}(9) = 3.2498$，故$\mu$的 99%置信区间为

$$\left[498.9 \pm 3.2498 \times \frac{4.38}{\sqrt{10}}\right] = [494.40, 503.40]\text{。}$$

三．估计方差σ^2

步骤：

（1）用样本方差 $S^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ 作为 σ^2 的点估计。



（2）建立枢轴量 $\chi^2=\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$。

（3）置信度 $1-\alpha$，即 $P\left\{\chi^2_{\alpha/2}(n-1)\le\dfrac{(n-1)S^2}{\sigma^2}\le\chi^2_{1-\alpha/2}(n-1)\right\}=1-\alpha$。

（4）解得 $\dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\le\sigma^2\le\dfrac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}$。

可得 σ^2 的 $1-\alpha$ 置信区间 $\left[\dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)},\dfrac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}\right]$。

例　设一批钢筋断强度服从 $N(\mu,\sigma^2)$，随取抽取 16 根，测得抗断强度为 289, 286, 285, 284, 286, 285, 285, 286, 292, 287, 281, 285, 291, 287, 284, 288，求 μ,σ^2 的 90%置信区间。

解：单个正态总体，未知 σ^2，估计 μ，用样本均值 $\bar{X}$ 作为 μ 的点估计，枢轴量 $T=\dfrac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$，则 μ 的 $1-\alpha$ 置信区间为 $\left[\bar{X}\pm t_{1-\alpha/2}(n-1)\cdot\dfrac{S}{\sqrt{n}}\right]$。

因 $\bar{x}=286.3$，$s^2=2.73^2$，$n=16$，$1-\alpha=0.9$，$t_{1-\alpha/2}(n-1)=t_{0.95}(15)=1.7531$，故 μ 的 90%置信区间为

$$\left[286.3\pm1.7531\times\frac{2.73}{\sqrt{16}}\right]=[285.1,287.5]。$$

估计 σ^2，用样本方差 S^2 作为 σ^2 的点估计，枢轴量 $\chi^2=\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$，则 σ^2 的 $1-\alpha$ 置信区间为 $\left[\dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)},\dfrac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}\right]$。

因 $s^2=2.73^2$，$n=16$，$1-\alpha=0.9$，$\chi^2_{\alpha/2}(n-1)=\chi^2_{0.05}(15)=7.2609$，$\chi^2_{1-\alpha/2}(n-1)=\chi^2_{0.95}(15)=24.9958$，故 σ^2 的 90%置信区间为

$$\left[\frac{15\times2.73^2}{24.9958},\frac{15\times2.73^2}{7.2609}\right]=[4.47,15.40]。$$

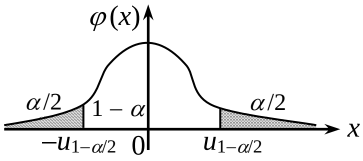
6.6.4　两个独立正态总体参数的置信区间

两个正态总体 $X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$，且相互独立。$X_1,X_2,\cdots,X_{n_1}$ 为来自总体 X 的样本，$Y_1,Y_2,\cdots,Y_{n_2}$ 为来自总体 Y 的样本，估计均值差 $\mu_1-\mu_2$ 或方差比 $\dfrac{\sigma_1^2}{\sigma_2^2}$。

一．已知方差 σ_1^2, σ_2^2，估计均值差 $\mu_1-\mu_2$

步骤：

（1）用样本均值之差 $\bar{X}-\bar{Y}$ 作为 $\mu_1-\mu_2$ 的点估计。



（2）建立枢轴量 $U=\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1)$。

（3）置信度 $1-\alpha$，即 $P\left\{-u_{1-\alpha/2}\le\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\le u_{1-\alpha/2}\right\}=1-\alpha$。

（4）解得 $\bar{X}-\bar{Y}-u_{1-\alpha/2}\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}\le\mu_1-\mu_2\le\bar{X}-\bar{Y}+u_{1-\alpha/2}\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}$。

可得 $\mu_1-\mu_2$ 的 $1-\alpha$ 置信区间 $\left[\bar{X}-\bar{Y}\pm u_{1-\alpha/2}\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}\right]$。

例　现有 A, B 两种型号的电子元件，A 型元件使用寿命 X（小时）服从正态分布 $N(\mu_1, 50^2)$，B 型元件使用寿命 Y（小时）服从正态分布 $N(\mu_2, 60^2)$。取 A 型元件 5 只，测得寿命为 1352, 1413, 1481, 1392, 1385（小时）；取 B 型元件 4 只，测得寿命为 980, 1100, 1052, 974（小时）。求两种元件寿命期望值之差 $\mu_1-\mu_2$ 的 98%置信区间。

解：两个独立正态总体，已知方差 σ_1^2, σ_2^2，估计 $\mu_1-\mu_2$，用样本均值之差 $\bar{X}-\bar{Y}$ 作为 $\mu_1-\mu_2$ 的点估计，枢轴量 $U=\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1)$，则 $\mu_1-\mu_2$ 的 $1-\alpha$ 置信区间为 $\left[\bar{X}-\bar{Y}\pm u_{1-\alpha/2}\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}\right]$。

因 $\bar{x}=1404.6$，$\bar{y}=1026.5$，$\sigma_1^2=50^2$，$\sigma_2^2=60^2$，$n_1=5$，$n_2=4$，$1-\alpha=0.98$，$u_{1-\alpha/2}=u_{0.99}=2.33$，

故 $\mu_1-\mu_2$ 的 98%置信区间为

$$\left[1404.6-1026.5\pm2.33\times\sqrt{\frac{50^2}{5}+\frac{60^2}{4}}\right]=[290.92, 465.28]。$$

二．未知方差 σ_1^2, σ_2^2，估计均值差 $\mu_1-\mu_2$

分四种情形讨论。

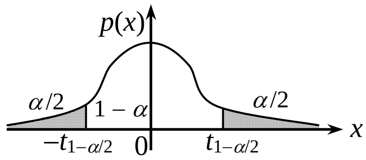
1．　方差 $\sigma_1^2=\sigma_2^2=\sigma^2$

已知 σ^2，有

$$U=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim N(0,1)\text{。}$$

未知σ^2，有

$$T=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)\text{，其中}S_w=\sqrt{\frac{(n_1-1)S_x^2+(n_2-1)S_y^2}{n_1+n_2-2}}\text{。}$$



步骤：

（1）用样本均值之差 $\bar{X}-\bar{Y}$ 作为 $\mu_1-\mu_2$ 的点估计。

（2）建立枢轴量 $T=\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$，其中 $S_w=\sqrt{\dfrac{(n_1-1)S_x^2+(n_2-1)S_y^2}{n_1+n_2-2}}$ 。

（3）置信度$1-\alpha$，即 $P\left\{-t_{1-\alpha/2}(n_1+n_2-2)\le\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\le t_{1-\alpha/2}(n_1+n_2-2)\right\}=1-\alpha$ 。

（4）解得 $\bar{X}-\bar{Y}-t_{1-\alpha/2}(n_1+n_2-2)\cdot S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\le\mu_1-\mu_2\le\bar{X}-\bar{Y}+t_{1-\alpha/2}(n_1+n_2-2)\cdot S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$ 。

可得 $\mu_1-\mu_2$ 的$1-\alpha$ 置信区间$\left[\bar{X}-\bar{Y}\pm t_{1-\alpha/2}(n_1+n_2-2)\cdot S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right]$， $S_w=\sqrt{\dfrac{(n_1-1)S_x^2+(n_2-1)S_y^2}{n_1+n_2-2}}$ 。

例　仍以前例中元件寿命为例，样本为 A 型：1352, 1413, 1481, 1392, 1385；B 型：980, 1100, 1052, 974。而方差未知（但 $\sigma_1^2=\sigma_2^2$），求寿命期望值之差 $\mu_1-\mu_2$ 的 98%置信区间。

解：两个独立正态总体，未知σ_1^2,σ_2^2，但$\sigma_1^2=\sigma_2^2$，估计 $\mu_1-\mu_2$，用样本均值之差 $\bar{X}-\bar{Y}$ 作为 $\mu_1-\mu_2$ 的点估计，枢轴量$T=\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$，则 $\mu_1-\mu_2$ 的$1-\alpha$ 置信区间为

$$\left[\bar{X}-\bar{Y}\pm t_{1-\alpha/2}(n_1+n_2-2)\cdot S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right]\text{，}\quad S_w=\sqrt{\frac{(n_1-1)S_x^2+(n_2-1)S_y^2}{n_1+n_2-2}}\text{。}$$

因 $\bar{x}=1404.6$，$\bar{y}=1026.5$，$s_x^2=48^2$，$s_y^2=60.47^2$，$n_1=5$，$n_2=4$，有 $S_w=\sqrt{\dfrac{4\times48^2+3\times60.47^2}{7}}=53.7$，

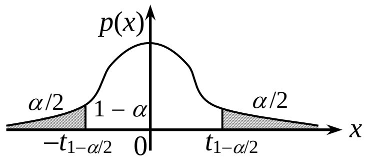
$1-\alpha=0.98$， $t_{1-\alpha/2}(n_1+n_2-2)=u_{0.99}(7)=2.9980$， 故 $\mu_1-\mu_2$ 的 98%置信区间为

$$\left[1404.6-1026.5\pm2.9980\times53.7\times\sqrt{\frac{1}{5}+\frac{1}{4}}\right]=[270.10,486.10]\text{。}$$

2．　方差 $\dfrac{\sigma_2^2}{\sigma_1^2}=c$ 为已知常数

这种情况下，有

$$T=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{c}{n_2}}}\sim t(n_1+n_2-2)\text{，其中 } S_w=\sqrt{\frac{(n_1-1)S_x^2+(n_2-1)S_y^2/c}{n_1+n_2-2}}\text{。}$$



步骤：

（1）用样本均值之差 $\bar{X}-\bar{Y}$ 作为 $\mu_1-\mu_2$ 的点估计。

（2）建立枢轴量 $T=\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{c}{n_2}}}\sim t(n_1+n_2-2)$，其中 $S_w=\sqrt{\dfrac{(n_1-1)S_x^2+(n_2-1)\dfrac{S_y^2}{c}}{n_1+n_2-2}}$。

（3）置信度$1-\alpha$，即 $P\left\{-t_{1-\alpha/2}(n_1+n_2-2)\le\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{c}{n_2}}}\le t_{1-\alpha/2}(n_1+n_2-2)\right\}=1-\alpha$。

（4）解得 $\bar{X}-\bar{Y}-t_{1-\alpha/2}(n_1+n_2-2)\cdot S_w\sqrt{\dfrac{1}{n_1}+\dfrac{c}{n_2}}\le\mu_1-\mu_2\le\bar{X}-\bar{Y}+t_{1-\alpha/2}(n_1+n_2-2)\cdot S_w\sqrt{\dfrac{1}{n_1}+\dfrac{c}{n_2}}$。

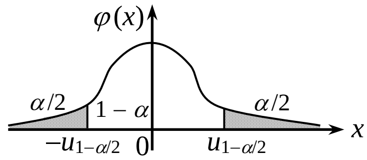
可得 $\mu_1-\mu_2$ 的$1-\alpha$ 置信区间 $\left[\bar{X}-\bar{Y}\pm t_{1-\alpha/2}(n_1+n_2-2)\cdot S_w\sqrt{\dfrac{1}{n_1}+\dfrac{c}{n_2}}\right]$，$S_w=\sqrt{\dfrac{(n_1-1)S_x^2+(n_2-1)\dfrac{S_y^2}{c}}{n_1+n_2-2}}$。

3．　未知二者关系，大样本场合

在未知 σ_1^2,σ_2^2 二者关系的情况下，一般就不能得到具有简单精确分布并且合适的枢轴量，此时根据 $U=\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1)$ 直接由替换原理得到 $\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_x^2}{n_1}+\dfrac{S_y^2}{n_2}}}$，其精确分布比较复杂，但在大样本场合近似服从标准正态分布 $N(0,1)$。

步骤：

（1）用样本均值之差 $\bar{X}-\bar{Y}$ 作为 $\mu_1-\mu_2$ 的点估计。



（2）建立枢轴量 $U=\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_x^2}{n_1}+\dfrac{S_y^2}{n_2}}}\mathrel{\dot\sim} N(0,1)$。

（3）置信度$1-\alpha$，即 $P\left\{-u_{1-\alpha/2}\le\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{S_x^2}{n_1}+\dfrac{S_y^2}{n_2}}}\le u_{1-\alpha/2}\right\}\approx1-\alpha$。

（4）解得 $\bar{X}-\bar{Y}-u_{1-\alpha/2}\sqrt{\frac{S_x^2}{n_1}+\frac{S_y^2}{n_2}}\le\mu_1-\mu_2\le\bar{X}-\bar{Y}+u_{1-\alpha/2}\sqrt{\frac{S_x^2}{n_1}+\frac{S_y^2}{n_2}}$ 。

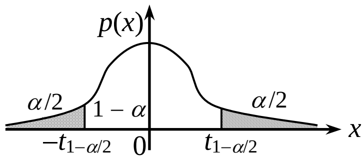
可得 $\mu_1-\mu_2$ 的 $1-\alpha$ 近似置信区间 $\left[\bar{X}-\bar{Y}\pm u_{1-\alpha/2}\sqrt{\frac{S_x^2}{n_1}+\frac{S_y^2}{n_2}}\right]$。

4． 未知二者关系，小样本场合

根据 $U=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)$ 直接由替换原理得到 $\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{S_x^2}{n_1}+\frac{S_y^2}{n_2}}}$，其精确分布比较复杂，

在小样本场合近似服从 t 分布 $t(l)$，自由度 $l\approx\frac{\left(\frac{S_x^2}{n_1}+\frac{S_y^2}{n_2}\right)^2}{\frac{S_x^4}{n_1^2(n_1-1)}+\frac{S_y^4}{n_2^2(n_2-1)}}$。

步骤：



（1）用样本均值之差 $\bar{X}-\bar{Y}$ 作为 $\mu_1-\mu_2$ 的点估计。

（2）建立枢轴量 $T=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{S_x^2}{n_1}+\frac{S_y^2}{n_2}}}\sim t(l)$，其中 $l\approx\frac{\left(\frac{S_x^2}{n_1}+\frac{S_y^2}{n_2}\right)^2}{\frac{S_x^4}{n_1^2(n_1-1)}+\frac{S_y^4}{n_2^2(n_2-1)}}$。

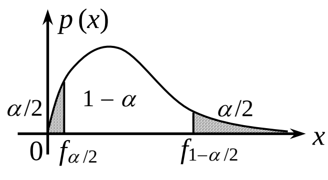
（3）置信度 $1-\alpha$，即 $P\left\{-t_{1-\alpha/2}(l)\le\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{S_x^2}{n_1}+\frac{S_y^2}{n_2}}}\le t_{1-\alpha/2}(l)\right\}=1-\alpha$ 。

（4）解得 $\bar{X}-\bar{Y}-t_{1-\alpha/2}(l)\cdot\sqrt{\frac{S_x^2}{n_1}+\frac{S_y^2}{n_2}}\le\mu_1-\mu_2\le\bar{X}-\bar{Y}+t_{1-\alpha/2}(l)\cdot\sqrt{\frac{S_x^2}{n_1}+\frac{S_y^2}{n_2}}$ 。

可得 $\mu_1-\mu_2$ 的 $1-\alpha$ 近似置信区间 $\left[\bar{X}-\bar{Y}\pm t_{1-\alpha/2}(l)\cdot\sqrt{\frac{S_x^2}{n_1}+\frac{S_y^2}{n_2}}\right]$，其中 $l\approx\frac{\left(\frac{S_x^2}{n_1}+\frac{S_y^2}{n_2}\right)^2}{\frac{S_x^4}{n_1^2(n_1-1)}+\frac{S_y^4}{n_2^2(n_2-1)}}$。

三．估计方差比 $\frac{\sigma_1^2}{\sigma_2^2}$

步骤：



（1）用样本方差之比 $\frac{S_x^2}{S_y^2}$ 作为 $\frac{\sigma_1^2}{\sigma_2^2}$ 的点估计。

（2）建立枢轴量 $F=\frac{S_x^2/\sigma_1^2}{S_y^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$ 。

（3）置信度$1-\alpha$，即$P\left\{f_{\alpha/2}(n_1-1,n_2-1)\le\frac{S_x^2/\sigma_1^2}{S_y^2/\sigma_2^2}\le f_{1-\alpha/2}(n_1-1,n_2-1)\right\}=1-\alpha$。

（4）解得$\frac{S_x^2}{S_y^2}\cdot\frac{1}{f_{1-\alpha/2}(n_1-1,n_2-1)}\le\frac{\sigma_1^2}{\sigma_2^2}\le\frac{S_x^2}{S_y^2}\cdot\frac{1}{f_{\alpha/2}(n_1-1,n_2-1)}$。

可得$\frac{\sigma_1^2}{\sigma_2^2}$的$1-\alpha$置信区间$\left[\frac{S_x^2}{S_y^2}\cdot\frac{1}{f_{1-\alpha/2}(n_1-1,n_2-1)},\frac{S_x^2}{S_y^2}\cdot\frac{1}{f_{\alpha/2}(n_1-1,n_2-1)}\right]$。

例　从总体$X\sim N(\mu_1,\sigma_1^2)$中取出容量为 15 的样本，测得$s_x^2=12.51^2$；从总体$Y\sim N(\mu_2,\sigma_2^2)$中取出容量为 26 的样本，测得$s_y^2=7.82^2$。求$\frac{\sigma_1^2}{\sigma_2^2}$的 90%置信区间。

解：两个正态总体，估计方差比$\frac{\sigma_1^2}{\sigma_2^2}$，用$\frac{S_x^2}{S_y^2}$作为$\frac{\sigma_1^2}{\sigma_2^2}$的点估计，枢轴量$F=\frac{S_x^2/\sigma_1^2}{S_y^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$，则$\frac{\sigma_1^2}{\sigma_2^2}$的$1-\alpha$置信区间为

$$\left[\frac{S_x^2}{S_y^2}\cdot\frac{1}{f_{1-\alpha/2}(n_1-1,n_2-1)},\frac{S_x^2}{S_y^2}\cdot\frac{1}{f_{\alpha/2}(n_1-1,n_2-1)}\right]。$$

因$s_x^2=12.51^2$，$s_y^2=7.82^2$，$n_1=15$，$n_2=26$，$1-\alpha=0.9$，$f_{1-\alpha/2}(n_1-1,n_2-1)=f_{0.95}(14,25)=2.11$，

$$f_{\alpha/2}(n_1-1,n_2-1)=f_{0.05}(14,25)=\frac{1}{f_{0.95}(25,14)}=\frac{1}{2.34}=0.4274,$$

故$\frac{\sigma_1^2}{\sigma_2^2}$的 90%置信区间为

$$\left[\frac{156.50}{61.15}\times\frac{1}{2.11},\frac{156.50}{61.15}\times2.34\right]=[1.213,5.988]。$$

6.6.5　　大样本场合概率p的置信区间

当样本容量$n\ge30$时，称之为大样本，在大样本场合下，可以用渐近分布构造近似的置信区间，很多情形下就用正态分布进行处理。

如在大样本场合，未知σ^2，估计μ，用$T=\frac{\bar X-\mu}{S/\sqrt{n}}\,\tilde{}\,N(0,1)$，可得$\mu$的$1-\alpha$近似置信区间为

$$\left[\bar X\pm u_{1-\alpha/2}\cdot\frac{S}{\sqrt{n}}\right]。$$

又如在大样本场合，估计事件A在一次试验中发生的概率p。设总体X为每一次试验中事件A的发生次数，总体X服从两点分布$b(1,p)$，$X_1,X_2,\cdots,X_n$为样本，以频率$\bar X=\frac{1}{n}\sum_{i=1}^{n}X_i$作为概率$p$的点估计。因

$\sum_{i=1}^{n} X_i$ 服从二项分布$b(n, p)$，根据棣莫弗-拉普拉斯中心极限定理知，当n很大时，$\frac{\sum_{i=1}^{n} X_i - np}{\sqrt{np(1-p)}} \sim N(0,1)$，

即$U = \frac{\bar{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0,1)$。置信度$1-\alpha$，有

$$P\left\{-u_{1-\alpha/2} \le \frac{\bar{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \le u_{1-\alpha/2}\right\} = 1-\alpha \text{。}$$

求解不等式

$$-u_{1-\alpha/2} \le \frac{\bar{X} - p}{\sqrt{\frac{p(1-p)}{n}}} \le u_{1-\alpha/2} \text{，}$$

但求解这个不等式与之前的不等式不同，不能简单地写成

$$\bar{X} - u_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} \le p \le \bar{X} + u_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}} \text{，}$$

可用以下三种方法处理这个不等式。

（1）近似法：将该不等式两端根号中的概率p换成频率$\bar{X}$，则可得p的$1-\alpha$近似置信区间为

$$\left[\bar{X} \pm u_{1-\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right] \text{。}$$

（2）方程法：由不等式

$$|\bar{X} - p| \le u_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

两边平方，并整理可得

$$\left(1 + \frac{u_{1-\alpha/2}^2}{n}\right)p^2 - \left(2\bar{X} + \frac{u_{1-\alpha/2}^2}{n}\right)p + \bar{X}^2 \le 0 \text{，}$$

方程的判别式

$$\Delta = \left(2\bar{X} + \frac{u_{1-\alpha/2}^2}{n}\right)^2 - 4\bar{X}^2\left(1 + \frac{u_{1-\alpha/2}^2}{n}\right) = \frac{4u_{1-\alpha/2}^2}{n}\bar{X}(1-\bar{X}) + \frac{u_{1-\alpha/2}^4}{n^2} > 0 \text{，}$$

记$\lambda = \frac{u_{1-\alpha/2}^2}{n}$，方程有两个相异实根

$$p_{\pm} = \frac{(2\bar{X} + \lambda) \pm \sqrt{\frac{4u_{1-\alpha/2}^2}{n}\bar{X}(1-\bar{X}) + \frac{u_{1-\alpha/2}^4}{n^2}}}{2(1+\lambda)} = \frac{\bar{X} + \frac{\lambda}{2}}{1+\lambda} \pm u_{1-\alpha/2}\frac{\sqrt{\frac{\bar{X}(1-\bar{X})}{n} + \frac{\lambda}{4n}}}{1+\lambda} \text{，}$$

可得p的$1-\alpha$近似置信区间为

$$\left[\frac{\bar{X}+\frac{\lambda}{2}}{1+\lambda} \pm u_{1-\alpha/2}\frac{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}+\frac{\lambda}{4n}}}{1+\lambda}\right],\quad \lambda=\frac{u_{1-\alpha/2}^2}{n}。$$

注： 当n很大时，如果对方程法的$1-\alpha$近似置信区间中的$\lambda=\frac{u_{1-\alpha/2}^2}{n}$近似取成 0，就得到近似法的结论。

（3）修正法：由于分位数$u_{1-\alpha/2}\approx 2$，有$\lambda\approx\frac{4}{n}$，则

$$\frac{\bar{X}+\frac{\lambda}{2}}{1+\lambda}\approx\frac{\bar{X}+\frac{2}{n}}{1+\frac{4}{n}}=\frac{\sum_{i=1}^{n}X_i+2}{n+4},$$

记修正频率$\bar{X}^*=\frac{\sum_{i=1}^{n}X_i+2}{n+4}$，并可以证明$\frac{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}+\frac{\lambda}{4n}}}{1+\lambda}\approx\sqrt{\frac{\bar{X}^*(1-\bar{X}^*)}{n+4}}$，则可得$p$的$1-\alpha$近似置信区间为

$$\left[\bar{X}^*\pm u_{1-\alpha/2}\sqrt{\frac{\bar{X}^*(1-\bar{X}^*)}{n+4}}\right]。$$

注： 修正法与近似法所得近似置信区间比较，只是将频率$\bar{X}=\frac{\sum_{i=1}^{n}X_i}{n}$换成修正频率$\bar{X}^*=\frac{\sum_{i=1}^{n}X_i+2}{n+4}$，样本容量$n$换成$n+4$，形式上类似，但近似程度要好得多。

例 从一批产品中抽取 100 件，发现有 80 件合格品，求合格率的 95%置信区间。

解： 大样本场合，估计概率p，用频率$\bar{X}$作为p的点估计，枢轴量$U=\frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}}\sim N(0,1)$。

近似法：因$\sum_{i=1}^{n}X_i=80$，$n=100$，频率$\bar{x}=\frac{80}{100}=0.8$，$1-\alpha=0.95$，$u_{1-\alpha/2}=u_{0.95}=1.96$，故$p$的 95%近似置信区间为

$$\left[\bar{X}\pm u_{1-\alpha/2}\sqrt{\frac{\bar{X}(1-\bar{X})}{n}}\right]=\left[0.8\pm 1.96\times\sqrt{\frac{0.8\times(1-0.8)}{100}}\right]=[0.7216, 0.8784]。$$

方程法：因$\sum_{i=1}^{n}X_i=80$，$n=100$，频率$\bar{x}=\frac{80}{100}=0.8$，$1-\alpha=0.95$，$u_{1-\alpha/2}=u_{0.95}=1.96$，故$p$的 95%近似置信区间为

$$\left[\frac{\bar{X}+\frac{\lambda}{2}}{1+\lambda}\pm u_{1-\alpha/2}\frac{\sqrt{\frac{\bar{X}(1-\bar{X})}{n}+\frac{\lambda}{4n}}}{1+\lambda}\right]=\left[\frac{0.8+\frac{1.96^2}{200}}{1+\frac{1.96^2}{100}}\pm 1.96\times\frac{\sqrt{\frac{0.8\times(1-0.8)}{100}+\frac{1.96^2}{4\times 100^2}}}{1+\frac{1.96^2}{100}}\right]$$

$$=[0.7112, 0.8666]。$$

修正法：因$\sum_{i=1}^{n}X_i=80$，$n=100$，修正频率$\bar{x}^*=\frac{82}{104}\approx 0.7885$，$1-\alpha=0.95$，$u_{1-\alpha/2}=u_{0.95}=1.96$，故

p 的 95%近似置信区间为

$$\left[\bar{X}^* \pm u_{1-\alpha/2}\sqrt{\frac{\bar{X}^*(1-\bar{X}^*)}{n+4}}\right] = \left[0.7885 \pm 1.96 \times \sqrt{\frac{0.7885\times(1-0.7885)}{104}}\right] = [0.7100, 0.8670]\text{。}$$

附表：

场合	枢轴量及其分布	置信区间
已知σ^2，估计μ	$U=\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$	$\left[\bar{X}\pm u_{1-\alpha/2}\dfrac{\sigma}{\sqrt{n}}\right]$
未知σ^2，估计μ	$T=\dfrac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$	$\left[\bar{X}\pm t_{1-\alpha/2}(n-1)\dfrac{S}{\sqrt{n}}\right]$
估计σ^2	$\chi^2=\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$	$\left[\dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)},\dfrac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}\right]$
已知σ_1^2,σ_2^2，估计$\mu_1-\mu_2$	$U=\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\sim N(0,1)$	$\left[\bar{X}-\bar{Y}\pm u_{1-\alpha/2}\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}\right]$
未知σ_1^2,σ_2^2，但$\sigma_1^2=\sigma_2^2$，估计$\mu_1-\mu_2$	$T=\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim t(n_1+n_2-2)$	$\left[\bar{X}-\bar{Y}\pm t_{1-\alpha/2}\cdot S_W\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}\right]$, $S_w=\sqrt{\dfrac{(n_1-1)S_x^2+(n_2-1)S_y^2}{n_1+n_2-2}}$
估计$\dfrac{\sigma_1^2}{\sigma_2^2}$	$F=\dfrac{S_x^2/\sigma_1^2}{S_y^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$	$\left[\dfrac{S_x^2}{S_y^2}\dfrac{1}{f_{1-\alpha/2}},\dfrac{S_x^2}{S_y^2}\dfrac{1}{f_{\alpha/2}}\right]$
大样本场合，估计概率p	$U=\dfrac{\bar{X}-p}{\sqrt{\dfrac{p(1-p)}{n}}}\sim N(0,1)$	$\left[\bar{X}\pm u_{1-\alpha/2}\sqrt{\dfrac{\bar{X}(1-\bar{X})}{n}}\right]$ $\left[\dfrac{\bar{X}+\dfrac{\lambda}{2}}{1+\lambda}\pm u_{1-\alpha/2}\dfrac{\sqrt{\dfrac{\bar{X}(1-\bar{X})}{n}+\dfrac{\lambda}{4n}}}{1+\lambda}\right]$ $\left[\bar{X}^*\pm u_{1-\alpha/2}\sqrt{\dfrac{\bar{X}^*(1-\bar{X}^*)}{n+4}}\right]$