第六章　参数估计

§6.1　点估计的概念与无偏性

1. 统计中的参数常指以下几种情况

- 分布中所含的未知参数 θ 及其某个函数 $g(\theta)$.
- 分布的各种特征数,如期望、方差、中位数等.

参数 θ 可能取值的范围 Θ 称为参数空间.

2. 参数估计的两种形式:点估计与区间估计

参数的点估计是指:对未知参数 θ 选用一个统计量 $\hat{\theta}=\hat{\theta}(x_1,x_2,\cdots,x_n)$ 的取值作为 θ 的估计值,$\hat{\theta}$就是 θ 的点估计(量),简称估计.好的点估计来自好的统计思想.区间估计见§6.6.

3. 无偏性与可估参数

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

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

则称$\hat{\theta}$是 θ 的无偏估计,否则称为有偏估计.

假如对任意 $\theta\in\Theta$,有$\lim\limits_{n\to\infty}E(\hat{\theta})=\theta$,则称$\hat{\theta}$是 θ 的渐近无偏估计.

并不是所有的参数都存在无偏估计,当参数存在无偏估计时称该参数是可估的.

4. 有效性　设$\hat{\theta}_1,\hat{\theta}_2$ 是 θ 的两个无偏估计,如果对任意的 $\theta\in\Theta$ 有

$$\mathrm{Var}(\hat{\theta}_1)\leqslant\mathrm{Var}(\hat{\theta}_2),$$

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

§6.2　矩估计及相合性

1. 矩法估计　利用如下两句“替换原理”而获得的估计量称为矩估计.

- 用样本矩替换总体矩,这里的矩可以是原点矩,也可以是中心矩.
- 用样本矩的函数去替换相应的总体矩的函数.

(1) 在总体分布未知场合,可用矩法对一些参数作出估计,如:

- 用样本均值 $\bar{x}$ 估计总体均值 $E(X)$.
- 用样本方差 s_n^2 估计总体方差 $\mathrm{Var}(X)$.
- 用事件 A 出现的频率估计事件 A 发生的概率 $P(A)$.
- 用样本分位数估计总体分位数.

(2) 在总体分布列或分布密度函数形式已知场合,在有关各阶矩存在的条件下,用“总体矩等于样本矩”列出矩方程(组),解之即得分布中未知参数的矩估计.其中尽量选用低阶矩.

2. 相合性　设 $\theta\in\Theta$ 为未知参数,$\hat{\theta}_n=\hat{\theta}_n(x_1,x_2,\cdots,x_n)$ 是 θ 的一个估计量,n 是样本容量,若对任何一个 $\varepsilon>0$,有

$$\lim_{n\to\infty}P(|\hat{\theta}_n-\theta|>\varepsilon)=0,\quad \forall\,\theta\in\Theta,$$

则称 $\hat{\theta}_n$ 为参数 θ 的相合估计.

相合性本质上就是按概率收敛,它是估计量的一个基本要求,即当样本量不断增大时,相合估计按概率收敛于未知参数.

矩法估计一般都是相合估计.

3. 判断相合性的一些定理

(1) 设 $\hat{\theta}_n=\hat{\theta}_n(x_1,x_2,\cdots,x_n)$ 是 θ 的一个估计量,若 $\lim\limits_{n\to\infty}E(\hat{\theta}_n)=\theta$, $\lim\limits_{n\to\infty}\mathrm{Var}(\hat{\theta}_n)=0$,则 $\hat{\theta}_n$ 是 θ 的相合估计.

(2) 若 $\hat{\theta}_{n1},\hat{\theta}_{n2},\cdots,\hat{\theta}_{nk}$ 分别是 $\theta_1,\theta_2,\cdots,\theta_k$ 的相合估计,$\eta=g(\theta_1,\theta_2,\cdots,\theta_k)$ 是 $\theta_1,\theta_2,\cdots,\theta_k$ 的连续函数,则 $\hat{\eta}_n=g(\hat{\theta}_{n1},\hat{\theta}_{n2},\cdots,\hat{\theta}_{nk})$ 是 η 的相合估计.

(3) 大数定律.

§6.3　最大似然估计与 EM 算法

1. 最大似然估计　利用“最大似然原理”获得的估计，它只能在总体概率函数形式已知的情况下使用，具体步骤如下：设总体的概率函数为 $p(x;\theta),\theta\in\Theta,x_1,x_2,\cdots,x_n$ 是来自该总体的样本.

- 写出似然函数 $L(\theta)=L(\theta;x_1,x_2,\cdots,x_n)=p(x_1;\theta)\cdot p(x_2;\theta)\cdot\cdots\cdot p(x_n;\theta)$.
- 使似然函数 $L(\theta)$ 达到最大的统计量 $\hat\theta=\hat\theta(x_1,x_2,\cdots,x_n)$ 称为 θ 的最大似然估计，简称 MLE，即 $\hat\theta$ 满足 $L(\hat\theta)=\max\limits_{\theta\in\Theta}L(\theta)$.

注意：使得对数似然函数 $\ln L(\theta)$ 最大的 $\hat\theta$ 也使似然函数 $L(\theta)$ 最大，寻找最大值可以从定义出发，也可以对 $l(\theta)=\ln L(\theta)$ 使用微分法，后者更为常用.

2. 最大似然估计的不变性　若 $\hat\theta$ 是 θ 的最大似然估计，则对任一函数 $g(\theta)$，$g(\hat\theta)$ 是其最大似然估计.

3. EM 算法　当分布中有多余参数或数据为截尾或缺失时，其 MLE 的求取是比较困难的.登普斯特(Dempster)等人于 1977 年提出了 EM 算法，其含义是把求 MLE 的过程分两步走，第一步求期望(E 步)，以便把多余的部分去掉，第二步求极大值(M 步).重复使用这两步直至收敛可得 MLE 的近似解.这是一种非常有效的方法.

4. MLE 的渐近正态性　在一般条件下，总体分布 $p(x;\theta)$ 中 θ 的 MLE $\hat\theta_n$ 具有相合性和渐近正态性，即 $\hat\theta_n\sim AN\left(\theta,\dfrac{1}{nI(\theta)}\right)$，其中 n 为样本容量，$I(\theta)=\int_{-\infty}^{\infty}\left(\dfrac{\partial\ln p}{\partial\theta}\right)^2p(x;\theta)\,\mathrm{d}x$ 为费希尔信息量.

§6.4　最小方差无偏估计

1. 均方误差　设 $\hat\theta$ 是 θ 的一个估计(无偏的或有偏的)，则称

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

为$\hat{\theta}$的均方误差.均方误差较小意味着:$\hat{\theta}$不仅方差较小,而且偏差$(E(\hat{\theta})-\theta)$也小,所以均方误差是评价点估计的最一般标准.

- 使均方误差一致最小的估计量一般是不存在的,但两个估计好坏可用均方误差评估.
- 在无偏估计类中使均方误差最小就是使方差最小.

2. 一致最小方差无偏估计 设$\hat{\theta}$是 θ 的一个无偏估计,如果对另外任意一个 θ 的无偏估计$\tilde{\theta}$,在参数空间 $\Theta=\{\theta\}$ 上都有

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

则称$\hat{\theta}$是 θ 的一致最小方差无偏估计,简记为 UMVUE.

3. 判断准则 设$\hat{\theta}=\hat{\theta}(x_1,x_2,\cdots,x_n)$是 θ 的一个无偏估计,$\mathrm{Var}(\hat{\theta})<+\infty$.如果对任意一个满足 $E(\varphi(x_1,x_2,\cdots,x_n))=0$ 的 φ,都有

$$\mathrm{Cov}_\theta(\hat{\theta},\varphi)=0,\quad \forall\,\theta\in\Theta,$$

则$\hat{\theta}$是 θ 的 UMVUE.

4. 充分性原则

- 任一参数 θ 的 UMVUE 不一定存在,若存在,则它一定是充分统计量的函数.
- 若 θ 的某个无偏估计$\hat{\theta}$不是充分统计量 $T=T(x_1,x_2,\cdots,x_n)$ 的函数,则通过条件期望可以获得一个新的无偏估计$\tilde{\theta}=E(\hat{\theta}\mid T)$,且方差比原估计的方差要小.
- 考虑 θ 的估计时,只需要在其充分统计量的函数中寻找即可,该说法对所有统计推断都是正确的.这便是充分性原则.

5. 费希尔信息量 $I(\theta)$ 设总体的概率函数 $p(x;\theta),\theta\in\Theta$ 满足下列条件:

(1) 参数空间 Θ 是直线上的一个开区间;

(2) 支撑 $S=\{x:p(x;\theta)>0\}$ 与 θ 无关;

(3) 导数$\frac{\partial}{\partial\theta}p(x;\theta)$对一切 $\theta\in\Theta$ 都存在;

(4) 对 $p(x;\theta)$，积分与微分运算可交换次序，即 $\frac{\partial}{\partial\theta}\int_{-\infty}^{+\infty} p(x;\theta)\,\mathrm{d}x = \int_{-\infty}^{+\infty} \frac{\partial}{\partial\theta} p(x;\theta)\,\mathrm{d}x$；

(5) 期望 $I(\theta) = E\left[\frac{\partial}{\partial\theta}\ln p(x;\theta)\right]^2$ 存在，

则称该期望 $I(\theta)$ 为总体分布的费希尔(Fisher)信息量.

如果二阶导数对一切 $\theta\in\Theta$ 都存在，则 $I(\theta)$ 还可用下式计算

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

6. 常用分布的费希尔信息量

- 二点分布 $b(1,p)$ 的费希尔信息量 $I(p)=[p(1-p)]^{-1}$.
- 泊松分布 $p(\lambda)$ 的费希尔信息量 $I(\lambda)=\lambda^{-1}$.
- 指数分布 $Exp(\lambda)$ 的费希尔信息量 $I(\lambda)=\lambda^2$.
- 正态分布 $N(\mu,1)$ 的费希尔信息量 $I(\mu)=1$.
- 正态分布 $N(0,\sigma^2)$ 的费希尔信息量 $I(\sigma^2)=\frac{1}{2\sigma^4}$.
- 正态分布 $N(\mu,\sigma^2)$ 的费希尔信息量(信息矩阵) $\boldsymbol{I}(\mu,\sigma^2)=\begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 1/(2\sigma^4) \end{pmatrix}$.

7. C-R 不等式

设 $T=T(x_1,x_2,\cdots,x_n)$ 是未知参数 $g(\theta)$ 的一个无偏估计，若 $g'(\theta)=\frac{\partial g(\theta)}{\partial\theta}$ 存在，则在费希尔信息量 $I(\theta)$ 也存在的条件下有

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

上式称为克拉默-拉奥(C-R)不等式，$[g'(\theta)]^2/(nI(\theta))$ 称为 $g(\theta)$ 的无偏估计的方差的 C-R 下界，简称 $g(\theta)$ 的 C-R 下界.特别，对 θ 的无偏估计 $\hat\theta$，有 $\mathrm{Var}(\hat\theta)\geqslant(nI(\theta))^{-1}$.

注：$g(\theta)$ 的 C-R 下界并不是对任意参数函数 $g(\theta)$ 的无偏估计的方差都可达到.但能达到 C-R 下界的 $g(\theta)$ 的估计 $T=T(x_1,x_2,\cdots,x_n)$ 一定是 $g(\theta)$ 的 UMVUE.方差达到 C-R 下界的无偏估计称为有效估计.

§ 6.5 贝叶斯估计

1. 贝叶斯统计推断使用的三种信息

- 总体信息，总体分布或总体所属分布族提供的信息.
- 样本信息，从总体中抽取样本所提供的信息.
- 先验信息，在试验前人们对要做的问题在经验上和资料上所拥有的信息.

2. 贝叶斯统计的基本观点

任一未知量 θ 都可看作一个随机变量，用一个概率分布来描述未知参数是最好的办法，这个分布称为先验分布.

3. 贝叶斯公式的密度函数形式

- 总体依赖于参数 θ 的概率函数在贝叶斯统计中记为 $p(x \mid \theta)$，它表示在随机变量 θ 取某个给定值时总体的条件概率函数.
- 根据参数 θ 的先验信息设法确定先验分布 $\pi(\theta)$.
- 从贝叶斯观点看，样本 $x_1,x_2,\cdots,x_n$ 的产生分两步进行.首先从先验分布$\pi(\theta)$产生一个样本 θ_0，然后从 $p(x_1,x_2,\cdots,x_n \mid \theta_0)$ 中产生一组样本.这时样本的**联合条件概率函数**为 $p(x_1,x_2,\cdots,x_n \mid \theta_0)=\prod\limits_{i=1}^{n} p(x_i \mid \theta_0)$，这个分布综合了总体信息和样本信息.
- θ_0 是未知的，它是按先验分布 $\pi(\theta)$ 产生的.为把先验信息综合进去，不能只考虑 θ_0，对 θ 的其他值发生的可能性也要加以考虑，故要用$\pi(\theta)$进行综合.这样一来，样本 $x_1,x_2,\cdots,x_n$ 和参数 θ 的联合分布为 $h(x_1,x_2,\cdots,x_n,\theta)=p(x_1,x_2,\cdots,x_n \mid \theta)\cdot\pi(\theta)$，这个联合分布把总体信息、样本信息和先验信息三种可用信息都综合进去了.
- 分析的目的是要对未知参数 θ 作统计推断.在没有样本信息时，人们只能依据先验分布对 θ 作出推断.在有了样本观察值 $x_1,x_2,\cdots,x_n$ 之后，则应依据$h(x_1,x_2,\cdots,x_n,\theta)$对 θ 作出推断.由于 $h(x_1,x_2,\cdots,x_n,\theta)$可分解为 $h(x_1,x_2,\cdots,x_n,\theta)=\pi(\theta \mid x_1,x_2,\cdots,x_n)m(x_1,x_2,\cdots,x_n)$，其中 $m(x_1,x_2,\cdots,$

$x_n) = \int_\Theta h(x_1, x_2, \cdots, x_n, \theta)\mathrm{d}\theta = \int_\Theta p(x_1, x_2, \cdots, x_n \mid \theta)\pi(\theta)\mathrm{d}\theta$ 是 $x_1, x_2, \cdots, x_n$ 的边际概率函数,它与 θ 无关,不含 θ 的任何信息.因此能用来对 θ 作出推断的仅是条件分布 $\pi(\theta \mid x_1, x_2, \cdots, x_n)$,它的计算公式是

$$\pi(\theta \mid x_1, x_2, \cdots, x_n) = \frac{h(x_1, x_2, \cdots, x_n, \theta)}{m(x_1, x_2, \cdots, x_n)} = \frac{p(x_1, x_2, \cdots, x_n \mid \theta)\pi(\theta)}{\int_\Theta p(x_1, x_2, \cdots, x_n \mid \theta)\pi(\theta)\mathrm{d}\theta},$$

这个条件分布称为 θ 的**后验分布**,它集中了总体、样本和先验中有关 θ 的一切信息.

后验分布 $\pi(\theta \mid x_1, x_2, \cdots, x_n)$ 的计算公式就是用密度函数表示的贝叶斯公式.它是用总体和样本对先验分布 $\pi(\theta)$ 作调整的结果,贝叶斯统计的一切推断都基于后验分布进行.

4. 贝叶斯估计 基于后验分布 $\pi(\theta \mid x_1, x_2, \cdots, x_n)$ 对 θ 所作的贝叶斯估计有多种,常用有如下三种:

- 使用后验分布的密度函数最大值作为 θ 的点估计,称为最大后验估计.
- 使用后验分布的中位数作为 θ 的点估计,称为后验中位数估计.
- 使用后验分布的均值作为 θ 的点估计,称为后验期望估计.这是使用最为频繁的贝叶斯估计.

5. 共轭先验分布 设 θ 是总体参数,$\pi(\theta)$ 是其先验分布,若对任意的样本观测值得到的后验分布 $\pi(\theta \mid \boldsymbol{X})$ 与 $\pi(\theta)$ 属于同一个分布族,则称该分布族是 θ 的共轭先验分布(族).

- 二项分布 $b(n, \theta)$ 中的成功概率 θ 的共轭先验分布是贝塔分布 $Be(a, b)$.
- 泊松分布 $P(\theta)$ 中的均值 θ 的共轭先验分布是伽马分布 $Ga(\alpha, \lambda)$.
- 在方差已知时,正态均值 θ 的共轭先验分布是正态分布 $N(\mu, \tau^2)$.
- 在均值已知时,正态方差 σ^2 的共轭先验分布是倒伽马分布 $IGa(\alpha, \lambda)$(若 $X \sim Ga(\alpha, \lambda)$,则 X^{-1} 的分布称为倒伽马分布 $IGa(\alpha, \lambda)$).

§6.6 区间估计

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

$$P_\theta(\hat{\theta}_L \leqslant \theta \leqslant \hat{\theta}_U) \geqslant 1-\alpha,$$

则称随机区间 $[\hat{\theta}_L, \hat{\theta}_U]$ 是 θ 的置信水平为 $1-\alpha$ 的置信区间,或简称 $[\hat{\theta}_L, \hat{\theta}_U]$ 是 θ 的 $1-\alpha$ 置信区间,$\hat{\theta}_L$ 和 $\hat{\theta}_U$ 分别称为 θ 的(双侧)置信下限和置信上限.

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

2. 同等置信区间 在上述记号下,若对给定的 $\alpha(0<\alpha<1)$,对任意的 $\theta \in \Theta$,有

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

则称 $[\hat{\theta}_L, \hat{\theta}_U]$ 为 θ 的 $1-\alpha$ 同等置信区间.

同等置信区间是把给定的置信水平 $1-\alpha$ 用足了.常在总体为连续分布场合下可以实现.

3. 置信限 在上述记号下,若对给定的 $\alpha(0<\alpha<1)$ 和任意的 $\theta \in \Theta$,有

$$P_\theta(\hat{\theta}_L \leqslant \theta) \geqslant 1-\alpha, \qquad \forall\, \theta \in \Theta,$$

则称 $\hat{\theta}_L$ 是 θ 的置信水平为 $1-\alpha$ 的(单侧)置信下限.假如等号对一切 $\theta \in \Theta$ 成立,则称 $\hat{\theta}_L$ 是 θ 的 $1-\alpha$ 同等置信下限.若对给定的 $\alpha(0<\alpha<1)$ 和任意的 $\theta \in \Theta$,有

$$P_\theta(\hat{\theta}_U \geqslant \theta) \geqslant 1-\alpha,$$

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

4. 枢轴量法 寻找同等置信区间常采用枢轴量法,其步骤如下:

- 设法构造一个样本和 θ 的函数 $G = G(x_1, x_2, \cdots, x_n, \theta)$,使得 G 的分布不依赖于未知参数.此种 G 被称为枢轴量.

- 适当地选择两个常数 c,d,使对给定的 $\alpha(0<\alpha<1)$,有 $P(c\leqslant G\leqslant d)=1-\alpha$.
- 若能将 $c\leqslant G\leqslant d$ 进行不等式等价变形化为 $\hat{\theta}_L\leqslant\theta\leqslant\hat{\theta}_U$,则有 $P_\theta(\hat{\theta}_L\leqslant\theta\leqslant\hat{\theta}_U)=1-\alpha$,最后的 $[\hat{\theta}_L,\hat{\theta}_U]$ 就是 θ 的 $1-\alpha$ 同等置信区间.

关于置信区间的构造有两点说明:

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

5. 常用的置信区间

(1) 设 $x_1,x_2,\cdots,x_n$ 是来自 $N(\mu,\sigma^2)$ 的样本,$\bar{x}$ 为样本均值,s 为样本标准差,u_p 为标准正态分布的 p 分位数,$t_p(k)$ 为自由度是 k 的 t 分布 $t(k)$ 的 p 分位数,$\chi_p^2(k)$ 为自由度是 k 的 χ^2 分布 $\chi^2(k)$ 的 p 分位数,取置信水平 $1-\alpha$,则

- σ 已知时 μ 的置信区间为 $[\bar{x}-u_{1-\alpha/2}\sigma/\sqrt{n},\bar{x}+u_{1-\alpha/2}\sigma/\sqrt{n}]$.
- σ 未知时 μ 的置信区间为 $[\bar{x}-t_{1-\alpha/2}s/\sqrt{n},\bar{x}+t_{1-\alpha/2}s/\sqrt{n}]$.
- σ^2(μ 未知)的置信区间为 $\left[\dfrac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)},\dfrac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)}\right]$.
- σ(μ 未知)的置信区间为 $\left[\dfrac{s\sqrt{n-1}}{\sqrt{\chi^2_{1-\alpha/2}(n-1)}},\dfrac{s\sqrt{n-1}}{\sqrt{\chi^2_{\alpha/2}(n-1)}}\right]$.

(2) 设 $x_1,x_2,\cdots,x_m$ 是来自 $N(\mu_1,\sigma_1^2)$ 的样本,$\bar{x}$ 为其样本均值,s_x 为其样本标准差,$y_1,y_2,\cdots,y_n$ 是来自 $N(\mu_2,\sigma_2^2)$ 的样本,$\bar{y}$ 为其样本均值,s_y 为其样本标准差,$u_p,t_p(k)$ 含义同上,$F_p(k_1,k_2)$ 为自由度是 (k_1,k_2) 的 F 分布 $F(k_1,k_2)$ 的 p 分位数,取置信水平 $1-\alpha$,则

• σ_1^2 与 σ_2^2 均已知时，$\mu_1-\mu_2$ 的置信区间为 $\left[\bar{x}-\bar{y}-u_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}},\bar{x}-\bar{y}+u_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}\right]$.

• $\sigma_1^2=\sigma_2^2$ 未知时，$\mu_1-\mu_2$ 的置信区间为$\left[\bar{x}-\bar{y}-\sqrt{\frac{m+n}{mn}}s_w t_{1-\alpha/2}(m+n-2),\bar{x}-\bar{y}+\sqrt{\frac{m+n}{mn}}s_w t_{1-\alpha/2}(m+n-2)\right]$，其中 $s_w^2=\frac{(m-1)s_x^2+(n-1)s_y^2}{m+n-2}$.

• $\frac{\sigma_1^2}{\sigma_2^2}=\theta$ 已知时，$\mu_1-\mu_2$ 的置信区间为$\left[\bar{x}-\bar{y}-\sqrt{\frac{m\theta+n}{mn}}s_t t_{1-\alpha/2}(m+n-2),\bar{x}-\bar{y}+\sqrt{\frac{m\theta+n}{mn}}s_t t_{1-\alpha/2}(m+n-2)\right]$，其中 $s_t^2=\frac{(m-1)s_x^2+(n-1)s_y^2/\theta}{m+n-2}$.

• m 与 n 都很大时，$\mu_1-\mu_2$ 的近似置信区间为 $\left[\bar{x}-\bar{y}-u_{1-\alpha/2}\sqrt{\frac{s_x^2}{m}+\frac{s_y^2}{n}},\bar{x}-\bar{y}+u_{1-\alpha/2}\sqrt{\frac{s_x^2}{m}+\frac{s_y^2}{n}}\right]$.

• 一般场合下 $\mu_1-\mu_2$ 的近似置信区间为 $[\bar{x}-\bar{y}-s_0t_{1-\alpha/2}(l),\bar{x}-\bar{y}+s_0t_{1-\alpha/2}(l)]$，其中 $s_0^2=s_x^2/m+s_y^2/n$，$l=\frac{s_0^4}{\frac{s_x^4}{m^2(m-1)}+\frac{s_y^4}{n^2(n-1)}}$.

• 方差比$\frac{\sigma_1^2}{\sigma_2^2}$的置信区间为$\left[\frac{s_x^2}{s_y^2}\cdot\frac{1}{F_{1-\alpha/2}(m-1,n-1)},\frac{s_x^2}{s_y^2}\cdot\frac{1}{F_{\alpha/2}(m-1,n-1)}\right]$.

（3）设 $x_1,x_2,\cdots,x_n$ 是来自 $b(1,p)$ 的样本，$\bar{x}$ 为其样本均值，则 n 很大时比例 p 的置信水平为 $1-\alpha$ 的近似置信区间为 $\left[\bar{x}-u_{1-\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}},\bar{x}+u_{1-\alpha/2}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right]$.

6. 样本量的确定 控制比率 p 的 $1-\alpha$ 置信区间长度不超过 $2d_0$ 的最小样本量为 $n \geqslant (u_{1-\alpha/2}/2d_0)^2$.