第 15 周　参数的区间估计

15.1 区间估计的基本思想

点估计用一个点(即一个数)估计未知参数。区间估计用一个区间估计未知参数。

例如估计一个人的年龄在 40 至 45 岁之间，一个人的身高在 1 米 75 至 1 米 80 之间，估计产品的合格率在 0.95 至 0.98 之间。

区间估计考虑到了估计的误差，给人们以更大的信任感。区间估计的理论就是用明确的概率语言刻画这种“信任感”的意义，并给出得到区间估计的具体概率方法。

****

例15.1.1 样本 X_1,X_2,X_3,X_4 来自正态总体 $N(\mu,1)$，样本均值 $\bar{X}$ 是 μ 的一个点估计，$[\bar{X}-1,\bar{X}+1]$ 为 μ 的一个区间估计。计算概率 $P(\bar{X}=\mu)$ 和 $P(\mu\in[\bar{X}-1,\bar{X}+1])$。

正态总体 $N(\mu,\sigma^2)$：

方差已知，期望未知（常见）　VS　期望已知，方差未知（罕见）

****

例15.1.1 样本 $X_1,\cdots,X_4$ 来自正态总体 $N(\mu,1)$，样本均值 $\bar{X}$ 是 μ 的一个点估计，$[\bar{X}-1,\bar{X}+1]$ 为 μ 的一个区间估计。计算 $P(\bar{X}=\mu)$ 和 $P(\mu\in[\bar{X}-1,\bar{X}+1])$。

解：$\bar{X}\sim N\left(\mu,\frac{1}{4}\right)$，$\frac{\bar{X}-\mu}{1/2}=2(\bar{X}-\mu)\sim N\left(0,\frac{1}{4}\right)$，　$P(\bar{X}=\mu)=0$

$$P(\mu\in[\bar{X}-1,\bar{X}+1])=P(\bar{X}-1\le\mu\le\bar{X}+1)$$

$$=P(\mu-1\le\bar{X}\le\mu+1)=P(-1\le\bar{X}-\mu\le1)$$

$$=P(-2\le2(\bar{X}-\mu)\le2)=\Phi(2)-\Phi(-2)\approx0.9544$$

****

区间估计的概率形式

设$I\left(X_1,X_2,\cdots,X_n\right)$是一个随机区间，

$\forall\ \theta\in\Theta$， $P_\theta\left(\theta\in I\left(X_1,X_2,\cdots,X_n\right)\right)\geq 1-\alpha$，

称该区间是参数θ的一个置信水平为$1-\alpha$（$0<\alpha<1$）的区间估计，

置信水平可以取到的最大值称为置信系数。

****

例15.1.1 样本$X_1,\cdots,X_4$ 来自正态总体$N\left(\mu,1\right)$，随机区间$\left[\bar{X}-1,\bar{X}+1\right]$为μ 的一个区间估计， $P\left(\mu\in\left[\bar{X}-1,\bar{X}+1\right]\right)=\Phi\left(2\right)-\Phi\left(-2\right)\approx 0.9544$

$P\left(\mu\in\left[\bar{X}-1,\bar{X}+1\right]\right)\approx 0.9544>0.8$

可以说$\left[\bar{X}-1,\bar{X}+1\right]$是μ 的一个置信水平0.8的区间估计，

$\left[\bar{X}-1,\bar{X}+1\right]$是μ 的一个置信水平0.9544的区间估计，置信系数为0.9544

****

区间估计的三种形式

$P_\theta\left(\theta\in I\left(X_1,X_2,\cdots,X_n\right)\right)\geq 1-\alpha$，

当$I=\left[\hat{\theta}_1\left(X_1,X_2,\cdots,X_n\right),\hat{\theta}_2\left(X_1,X_2,\cdots,X_n\right)\right]$时，称$I$之为$\theta$的置信水平为$1-\alpha$（双侧）置信区间，

当$I=\left[\hat{\theta}_L\left(X_1,X_2,\cdots,X_n\right),+\infty\right]$时，称$I$之为$\theta$ 的置信水平为$1-\alpha$单侧置信区间，

$\hat{\theta}_L$为θ 的置信水平为$1-\alpha$置信下限，

当$I=\left[-\infty,\hat{\theta}_U\left(X_1,X_2,\cdots,X_n\right)\right]$时，称$I$之为$\theta$的置信水平为$1-\alpha$单侧置信区间，

$\hat{\theta}_U$为 θ 的置信水平为$1-\alpha$置信上限。

****

例15.1.1 样本$X_1,\cdots,X_4$ 来自正态总体$N(\mu,1)$，随机区间$\left[\bar{X}-1,\bar{X}+1\right]$是$\mu$ 的一个置信水平0.9544的区间估计。

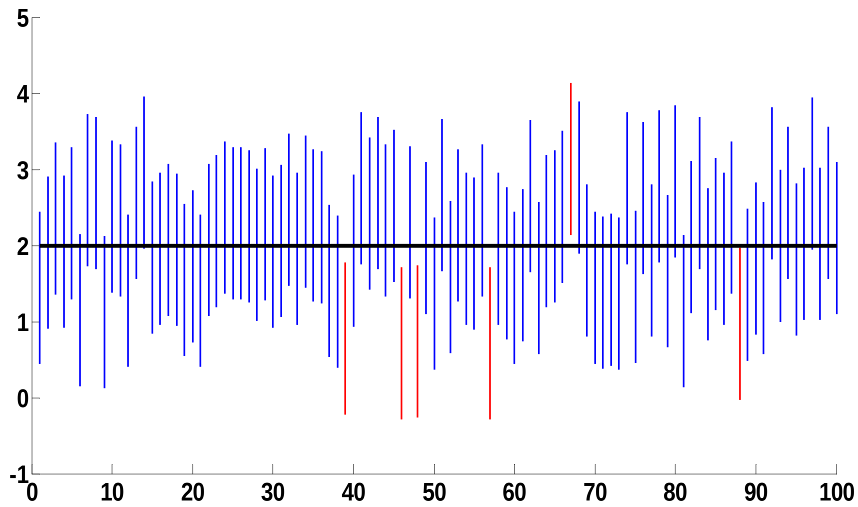

图 15.1

事实上，未知参数本身是确定的值，不带有随机性。估计的随机性是由区间引入的。一个置信水平$1-\alpha$的区间估计，其含义是：所得到的随机区间至少以概率$1-\alpha$覆盖被估的参数。在例 15.1.1 中，置信水平的含义是：实际抽取样本容量为 4 的样本得到一次区间估计，将这样的估计重复足够多次，那么至少$1-\alpha$比例的估计区间包含真实的μ 值。这里置信系数是 0.9544，即将这一估计方法实际重复足够多次后，大约 95.44%的估计区间包含真实的μ 值。图 15.1 是设定参数 mu 等于 2，将区间估计重复 100 次的模拟结果，其中 100 个估计区间中有 6 个区间没有包含 2，即参数的实际取值，94%的区间成功的包含了真实值 mu 等于 2。继续模拟，将区间估计重复 10000 次，结果 442 次估计区间没有包含 mu 的真实值 2，这时成功包含被估参数 mu 等于 2 的区间百分比达到 95.58%。当然，这些都是概率意义上的结果。具体的每一次估计，我们不会知道区间是否包含未知参数。

****

15.2 区间估计的构造方法

区间估计的两个基本要求：

1．未知参数θ要以尽可能大的概率落在区间 $I(X_1, X_2, \cdots, X_n)$ 中；

2. 在达到一定的置信水平的前提下，要求区间的长度尽可能小。

****

参数区间估计枢轴量法实现步骤

步骤 1　构造“枢轴量”（pivot）$G(X_1, X_2, \cdots, X_n, \theta)$，$G$ 的分布不依赖于任何未知参数；

步骤 2　适当选取两个常数 c、d，对给定的 α（$0<\alpha<1$），有

$$P(c \le G \le d) \ge 1-\alpha$$

步骤 3　求解 $c \le G(X_1, X_2, \cdots, X_n, \theta) \le d$，得到

$$\hat{\theta}_1(X_1, X_2, \cdots, X_n) \le \theta \le \hat{\theta}_2(X_1, X_2, \cdots, X_n)。$$

****

例 15.2.1　总体 $N(\mu, \sigma^2)$，σ^2 已知，μ 未知，简单随机样本 $X_1, X_2, \cdots, X_n$，求 μ 的 $1-\alpha$ 置信区间。

解：被估参数 μ 的一个点估计为 $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$，

$$G = \frac{\bar{X}-\mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim N(0,1)$$

是样本和被估参数 μ 的函数，且 G 的分布不依赖任何未知参数，可作为枢轴量。

取适当的常数 c 和 d，使得 $\Phi(d)-\Phi(c) = P\left(c \le \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \le d\right) \ge 1-\alpha$，

取 $d = u_{1-\alpha/2}$，$c = u_{\alpha/2} = -u_{1-\alpha/2}$

即得到参数μ 的$1-\alpha$置信区间 $\bar{X}-u_{1-\alpha/2}\frac{\sqrt{n}}{\sigma}\le\mu\le\bar{X}+u_{1-\alpha/2}\frac{\sqrt{n}}{\sigma}$。

****

枢轴量不等式 $P(c\le G\le d)\ge 1-\alpha$ 中c 和d 的取法

枢轴量为对称分布时

c 和d 的最优值为G的分布函数的$\frac{\alpha}{2}$和$1-\frac{\alpha}{2}$分位点。

当枢轴量的分布为非对称分布时

最优的c 和d 可能不易求得，通常也取枢轴量分布的$\frac{\alpha}{2}$和$1-\frac{\alpha}{2}$分位数点。

如在例 15.2.1 中c 和d 选取的是最优值，分别为$d=u_{1-\frac{\alpha}{2}}$，$c=-d$。

****

回顾统计抽样定理 设$X_1,X_2,\cdots,X_n$ 是来自正态总体$N(\mu,\sigma^2)$ 的样本，其样本均值和样本方差分别为$\bar{X}$和S^2，则有

（1）$\bar{X}$和S^2相互独立，

（2）$\bar{X}\sim N\left(\mu,\frac{\sigma^2}{n}\right)$，

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

我们在下面构造区间估计的枢轴量的时候，要用到这几个结论。

****

例 15.2.2 总体 $N(\mu,\sigma^2)$，μ 未知，σ^2未知，$X_1,X_2,\cdots,X_n$是来自该总体的样本，求σ^2的$1-\alpha$置信区间。

解； 正态总体的样本方差满足$\frac{(n-1)\cdot S^2}{\sigma^2}\sim\chi^2(n-1)$

构造枢轴量 $G=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$，

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

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

****

例 15.2.3 总体 $N(\mu,\sigma^2)$，μ 未知，σ^2 未知，$X_1,X_2,\cdots,X_n$是来自该总体的样本，求μ 的$1-\alpha$置信区间。

分析： $\bar{X}\sim N\left(\mu,\dfrac{\sigma^2}{n}\right)$， $\dfrac{\bar{X}-\mu}{\sqrt{\dfrac{\sigma^2}{n}}}=\dfrac{\sqrt{n}\left(\bar{X}-\mu\right)}{\sigma}\sim N(0,1)$

σ未知，所以这个标准正态分布随机变量不能作为枢轴量

解： 同时考虑 $\dfrac{\sqrt{n}\left(\bar{X}-\mu\right)}{\sigma}\sim N(0,1)$ ， $\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$

正态总体的样本均值$\bar{X}$与样本方差S^2相互独立

构造 t 统计量， $\dfrac{\dfrac{\sqrt{n}\left(\bar{X}-\mu\right)}{\sigma}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2}\Big/ n-1}}=\dfrac{\sqrt{n}\left(\bar{X}-\mu\right)}{S}\sim t(n-1)$

由于 $P\left(t_{\alpha/2}(n-1)\le\dfrac{\sqrt{n}\left(\bar{X}-\mu\right)}{S}\le t_{1-\alpha/2}(n-1)\right)=1-\alpha$

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

****

例 15.2.4 假设工厂生产的产品的质量服从正态分布，估计所生产的产品的平均

质量，现从所生产的产品中抽取9件，用一台天平测量，测得结果分别为（单位克）7.32，7.37，7.24，7.41，7.33，7.38，7.36，7.42，7.21。

试求产品平均质量的0.95置信区间。

解： 假设产品的质量服从正态分布$N(\mu,\sigma^2)$，这里μ未知，σ^2未知，

$$\frac{\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma}}{\sqrt{\frac{(n-1)S^2}{\sigma^2}\Big/n-1}}=\frac{\sqrt{n}(\bar{X}-\mu)}{S}\sim t(n-1),\quad \text{（格赛特，student）}$$

μ的$1-\alpha$置信区间$\left[\bar{X}-\frac{S}{\sqrt{n}}t_{1-\alpha/2}(n-1),\bar{X}+\frac{S}{\sqrt{n}}t_{1-\alpha/2}(n-1)\right]$

计算此次抽样样本均值与样本方差的取值

$$\bar{x}=\frac{7.32+7.37+7.24+7.41+7.33+7.38+7.36+7.42+7.21}{9}=7.338,$$

$$s^2=\frac{1}{8}\sum_{i=1}^{9}(x_i-\bar{x})^2=5.19\times10^{-3},\quad s=\sqrt{5.19\times10^{-3}}=0.072。$$

查表得$t_{0.975}(8)=2.306$，得到μ的置信系数0.95的区间估计为$[7.282,7.393]$。

****

15.3 两个正态总体的区间估计

两个正态总体$X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$，$\mu_1$、$\mu_2$未知，$\sigma_1^2$、$\sigma_2^2$未知，$X_1,\cdots,X_m$和$Y_1,\cdots,Y_n$分别是来自两个总体的样本，且假设这两个样本相互独立，求$\mu_2-\mu_1$的$1-\alpha$置信区间。

此问题被命名为贝伦斯-费舍尔（Behrens-Fisher）问题，是统计学中至今没有完全解决的问题。

****

例15.3.1 两个正态总体$X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$，$\mu_1$、$\mu_2$未知，$\sigma_1^2$、$\sigma_2^2$已

知， $X_1,\cdots,X_m$ 和 $Y_1,\cdots,Y_n$ 分别是来自两个总体的样本，且假设这两个样本相互独立，求 $\mu_2-\mu_1$ 的 $1-\alpha$ 置信区间。

解： $\bar{X}\sim N\left(\mu_1,\frac{\sigma_1^2}{m}\right)$， $\bar{Y}\sim N\left(\mu_2,\frac{\sigma_2^2}{n}\right)\quad\Rightarrow\quad\bar{X}-\bar{Y}\sim N\left(\mu_1-\mu_2,\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}\right)$

$$G=\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}\sim N(0,1)，\qquad P\left(-u_{1-\alpha/2}\le\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}\le u_{1-\alpha/2}\right)=1-\alpha$$

$\mu_2-\mu_1$ 的 $1-\alpha$ 置信区间 $\left[\bar{X}-\bar{Y}-u_{1-\alpha/2}\cdot\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}},\bar{X}-\bar{Y}+u_{1-\alpha/2}\cdot\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}\right]$

****

例 15.3.2 两个正态总体 $X\sim N(\mu_1,\sigma_1^2)$， $Y\sim N(\mu_2,\sigma_2^2)$， μ_1、 μ_2 未知， σ_1^2、 σ_2^2 未知但相等，即 $\sigma_1^2=\sigma_2^2=\sigma^2$， $X_1,\cdots,X_m$ 和 $Y_1,\cdots,Y_n$ 分别是来自两个总体的样本，且假设这两个样本相互独立，求 $\mu_2-\mu_1$ 的 $1-\alpha$ 置信区间。

解： $\bar{X}\sim N\left(\mu_1,\frac{\sigma^2}{m}\right)$， $\bar{Y}\sim N\left(\mu_2,\frac{\sigma^2}{n}\right)$， $\bar{X}-\bar{Y}\sim N\left(\mu_1-\mu_2,\left(\frac{1}{m}+\frac{1}{n}\right)\sigma^2\right)$，

标准化后 $\dfrac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sigma\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right)}}\sim N(0,1)$，

$$\frac{(m-1)S_x^2}{\sigma^2}\sim\chi^2(m-1)，\quad\frac{(n-1)S_y^2}{\sigma^2}\sim\chi^2(n-1)，\quad\frac{(m-1)S_x^2+(n-1)S_y^2}{\sigma^2}\sim\chi^2(m+n-2)$$

$$\frac{\left.\bar{X}-\bar{Y}-(\mu_1-\mu_2)\middle/\sigma\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right)}\right.}{\sqrt{\left.\frac{(m-1)S_x^2+(n-1)S_y^2}{\sigma^2}\middle/m+n-2\right.}}=\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right)}\sqrt{\frac{(m-1)S_x^2+(n-1)S_y^2}{m+n-2}}}\sim t(m+n-2)$$

记$S_w^2=\dfrac{(m-1)S_x^2+(n-1)S_y^2}{m+n-2}$，

$$P\left(-t_{1-\alpha/2}(m+n-2)\le\frac{\bar X-\bar Y-(\mu_1-\mu_2)}{S_w\sqrt{\left(\frac{1}{m}+\frac{1}{n}\right)}}\le t_{1-\alpha/2}(m+n-2)\right)=1-\alpha$$

得到$\mu_2-\mu_1$的$1-\alpha$置信区间

$$\left[\bar X-\bar Y-\sqrt{\frac{1}{m}+\frac{1}{n}}S_w t_{1-\alpha/2}(m+n-2),\bar X-\bar Y+\sqrt{\frac{1}{m}+\frac{1}{n}}S_w t_{1-\alpha/2}(m+n-2)\right]。$$

****

例 15.3.3　两个正态总体$X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$，$\mu_1$、$\mu_2$未知，$\sigma_1^2$、$\sigma_2^2$未知，$X_1,\cdots,X_m$和$Y_1,\cdots,Y_n$分别是来自两个总体的样本,假设这两个样本相互独立,求σ_1^2/σ_2^2的$1-\alpha$置信区间。

解：$\dfrac{(m-1)S_x^2}{\sigma_1^2}\sim\chi^2(m-1)$ 和 $\dfrac{(n-1)S_y^2}{\sigma_2^2}\sim\chi^2(n-1)$，两者相互独立,

故构造F分布枢轴量 $F=\dfrac{S_x^2/\sigma_1^2}{S_y^2/\sigma_2^2}\sim F(m-1,n-1)$,

σ_1^2/σ_2^2的$1-\alpha$置信区间 $\left[\dfrac{S_x^2}{S_y^2}\dfrac{1}{F_{1-\alpha/2}(m-1,n-1)},\dfrac{S_x^2}{S_y^2}\dfrac{1}{F_{\alpha/2}(m-1,n-1)}\right]$。

****

15.4 大样本置信区间

样本容量足够大时，根据中心极限定理可以利用渐近正态分布构造置信区间。

例 15.4.1 贝伦斯-费舍尔（Behrens-Fisher）问题，样本$X_1,\cdots,X_m$来自正态总体 $N(\mu_1,\sigma_1^2)$，样本$Y_1,\cdots,Y_n$来自正态总体 $N(\mu_2,\sigma_2^2)$，μ_1、μ_2未知，σ_1^2、σ_2^2

未知，且假设这两个样本相互独立，求$\mu_2-\mu_1$的$1-\alpha$置信区间。

解： $\bar{X}\sim N\left(\mu_1,\frac{\sigma_1{}^2}{m}\right)$， $\bar{Y}\sim N\left(\mu_2,\frac{\sigma_2{}^2}{n}\right)$ $\Rightarrow$ $\bar{X}-\bar{Y}\sim N\left(\mu_1-\mu_2,\frac{\sigma_1{}^2}{m}+\frac{\sigma_2{}^2}{n}\right)$

$$\frac{\bar{X}-\bar{Y}-(\mu_2-\mu_1)}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}\sim N(0,1)$$

当 m 和 n 较大时， $S_x^2\approx\sigma_1{}^2$ $S_y^2\approx\sigma_2{}^2$

有如下近似 $\dfrac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{S_x^2/m+S_y^2/n}}\sim N(0,1)$，以此作为枢轴量，

可得到$\mu_2-\mu_1$的$1-\alpha$置信水平的

近似区间估计$\left[\bar{X}-\bar{Y}-u_{1-\alpha/2}\sqrt{S_x^2/m+S_y^2/n},\ \bar{X}-\bar{Y}+u_{1-\alpha/2}\sqrt{S_x^2/m+S_y^2/n}\right]$。

****

例 15.4.2 样本$X_1,\cdots,X_n$来自两点分布总体$b(1,p)$，求 p 的区间估计。

解：样本均值的期望、方差分别为 $E(\bar{X})=p$，$Var(\bar{X})=\dfrac{p(1-p)}{n}$

根据中心极限定理当n较大时，有近似分布

$\bar{X}\sim N\left(p,\dfrac{p(1-p)}{n}\right)$， 标准化后得到枢轴量 $\dfrac{\bar{X}-p}{\sqrt{\dfrac{p(1-p)}{n}}}\sim N(0,1)$，

$P\left(\left|\dfrac{\bar{X}-p}{\sqrt{p(1-p)/n}}\right|\le u_{1-\frac{\alpha}{2}}\right)\approx 1-\alpha$， 即 $(\bar{X}-p)^2\le u_{1-\frac{\alpha}{2}}^2\dfrac{p(1-p)}{n}$

$P\left(\left|\dfrac{\bar{X}-p}{\sqrt{p(1-p)/n}}\right|\le u_{1-\frac{\alpha}{2}}\right)\approx 1-\alpha$， 即 $(\bar{X}-p)^2\le u_{1-\frac{\alpha}{2}}^2\dfrac{p(1-p)}{n}$，解得

$$\frac{1}{1+c}\left(\bar{X}+\frac{c}{2}-\sqrt{\frac{\bar{X}(1-\bar{X})}{n}u_{1-\frac{\alpha}{2}}^2+\frac{c^2}{4}}\right)\le p\le\frac{1}{1+c}\left(\bar{X}+\frac{c}{2}+\sqrt{\frac{\bar{X}(1-\bar{X})}{n}u_{1-\frac{\alpha}{2}}^2+\frac{c^2}{4}}\right)$$

其中 $c=\frac{u_{1-\frac{\alpha}{2}}^2}{n}$， 当$n$较大时，$c$的值很小可略去，得到参数 p 的$1-\alpha$置信水平的近似估计区间$\left[\bar{X}-u_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}\left(1-\bar{X}\right)}{n}},\bar{X}+u_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}\left(1-\bar{X}\right)}{n}}\right]$。

****

例 15.4.3 一个网络产品的运营商希望了解该产品在某个城市的用户占有率，进行了随机抽样调查，样本容量为 500，调查结果有 84 人是该产品的用户，求这个网络产品在该城市占有率的一个 95%置信区间。

解：总体分布 $X\sim b(1,p)$， 样本均值 $\bar{X}=\frac{X_1+\cdots+X_{500}}{500}$

期望、方差分别为 $E\left(\bar{X}\right)=p$，$Var\left(\bar{X}\right)=\frac{p(1-p)}{n}=\frac{p(1-p)}{500}$

$\frac{\left(\bar{X}-p\right)}{\sqrt{p(1-p)/500}}\sim N(0,1)$， $\left[\bar{X}-u_{0.975}\sqrt{\frac{\bar{X}\left(1-\bar{X}\right)}{500}},\bar{X}+u_{0.975}\sqrt{\frac{\bar{X}\left(1-\bar{X}\right)}{500}}\right]$

$\bar{x}=\frac{84}{500}=0.168$， $u_{0.975}=1.96$， 近似估计区间为：$[0.135,0.201]$。

****

例 15.4.4 一个网络产品的运营商希望了解该产品在某个城市的用户占有率，进行了随机抽样调查，样本容量为 500，调查结果有 84 人是该产品的用户。根据抽样调查的信息，运营商希望得到参数 p 的 95%置信水平的区间估计，且估计区间长度不超过 0.02，问至少需要多大的样本容量？

解：参数 p 的$1-\alpha$置信区间为$\left[\bar{X}-u_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}\left(1-\bar{X}\right)}{n}},\bar{X}+u_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}\left(1-\bar{X}\right)}{n}}\right]$

估计区间长度不超过 0.02，即$u_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{x}\left(1-\bar{x}\right)}{n}}\le 0.01$。

$$n \geq \left(\frac{u_{1-\alpha/2}}{0.01}\right)^2 \bar{x}\left(1-\bar{x}\right) = \left(\frac{u_{0.975}}{0.01}\right)^2 \bar{x}\left(1-\bar{x}\right) = \left(\frac{1.96}{0.01}\right)^2 0.168\left(1-0.168\right) = 5369.6$$

****

例 15.4.5 一个网络产品的运营商希望了解该产品在某个城市的用户占有率，希望得到参数 p 的 95%置信水平，且估计区间长度不超过 0.02 的区间估计，问在没有任何先验知识的情况下，至少需要多大的样本容量才能保证达到所希望的估计精度？

解：参数 p 的 $1-\alpha$ 置信区间为 $\left[\bar{X} - u_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}\left(1-\bar{X}\right)}{n}}, \bar{X} + u_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}\left(1-\bar{X}\right)}{n}}\right]$

对任何 $\bar{X} \in \left(0,1\right)$，$\bar{x}\left(1-\bar{x}\right) \leq \frac{1}{4}$

$$n \geq \left(\frac{u_{1-\alpha/2}}{0.01}\right)^2 \frac{1}{4} \geq \left(\frac{u_{1-\alpha/2}}{0.01}\right)^2 \bar{x}\left(1-\bar{x}\right), \qquad n \geq \left(\frac{u_{0.975}}{0.01}\sqrt{\frac{1}{4}}\right)^2 = 9604$$

问题没有变，可是现在这个情况下估计的所需样本容量数比上一题多了不少。因为本题中先验信息少于上一题，所以得到估计不如上题精确也是很自然的。一般而言，得到信息越多，越有可能得到更好的估计。

****