数理统计

上海财经大学　统计与管理学院

Contents

第三章 参数估计

§3.1 点估计的概念与无偏性

- 一般常用θ表示参数，参数θ所有可能取值组成的集合称为**参数空间**，常用Θ表示。参数估计问题就是根据样本对上述各种未知参数作出估计。

- 参数估计的形式有两种：**点估计与区间估计。**

3.1.1 点估计及无偏性

- **定义3.1** 设$X_1, X_2, \dots, X_n$是来自总体X的一个样本，我们用一个统计量$\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$的取值作为$\theta$的估计值，$\hat{\theta}$称为$\theta$的**点估计（量）**，简称**估计**。在这里如何构造统计量$\hat{\theta}$并没有明确的规定，只要它满足一定的合理性即可。这就涉及到两个问题：
  - 其一 是如何给出估计，即估计的方法问题；
  - 其二 是如何对不同的估计进行评价，即估计的好坏判断标准。

3.1.1 点估计及无偏性

- 定义3.2 设$\hat{\theta}=\hat{\theta}(X_1,X_2,\ldots,X_n)$是$\theta$的一个估计，$\theta$的参数空间为$\Theta$，若对任意的$\theta\in\Theta$，有

$$E_\theta\left(\hat{\theta}\right)=\theta$$

则称$\hat{\theta}$是θ的无偏估计，否则称为有偏估计。

3.1.1 点估计及无偏性

- 补1 设$X_1, X_2, \dots, X_n$为抽自均值为μ的总体的样本，考虑μ的估计量

$$\hat{\mu}_1 = X_1,$$
$$\hat{\mu}_2 = \frac{X_1 + X_2}{2},$$
$$\hat{\mu}_3 = \frac{X_1 + X_2 + X_{n-1} + X_n}{4} \quad (\text{假设} n \geq 4).$$

因为$E(X_i) = \mu$，容易验证，$E(\hat{\mu}_i) = \mu, i = 1,2,3$，所以$\hat{\mu}_1, \hat{\mu}_2$和$\hat{\mu}_3$都是$\mu$的无偏估计。但是

$$\hat{\mu}_4 = 2X_1,$$
$$\hat{\mu}_5 = \frac{X_1 + X_2}{3}$$

都不是μ的无偏估计。

3.1.1 点估计及无偏性

- 例3.1 对任一总体而言，样本均值是总体均值的无偏估计。当总体k阶矩存在时，样本k阶原点矩A_k是总体k阶原点矩μ_k的无偏估计。但对k阶中心矩则不一样，譬如，由于$E(S_n^2)=\frac{n-1}{n}\sigma^2$，样本方差$S_n^2$不是总体方差$\sigma^2$的无偏估计，对此，有如下两点说明：

  (1) 当样本量趋于无穷时，有$E(S_n^2)\to\sigma^2$，我们称S_n^2为σ^2的**渐近无偏估计。**

  (2) 若对S_n^2作如下修正：

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

  则S^2是总体方差的无偏估计。

3.1.1 点估计及无偏性

- 例3.2 设总体为$N(\mu,\sigma^2)$，$X_1,X_2,\dots,X_n$是样本，则S^2是σ^2的无偏估计，且可求出

$$E(S)=\sqrt{\frac{2}{n-1}}\cdot\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}\cdot\sigma\equiv\frac{\sigma}{c_n}$$

这说明S不是σ的无偏估计。

利用修正技术可得c_nS是σ的无偏估计，其中

$$c_n=\sqrt{\frac{n-1}{2}}\cdot\frac{\Gamma((n-1)/2)}{\Gamma(n/2)}$$是修偏系数。

可以证明，当$n\to\infty$时, 有$c_n\to 1$。说明S是σ的渐近无偏估计。

3.1.1 点估计及无偏性

- **补2** 设$X_1,X_2,\dots,X_n$是来自$Exp(\lambda)$的样本，已知$\bar{X}$为$1/\lambda$的无偏估计，试说明$1/\bar{X}$是否为λ的无偏估计。
- **解：** 因为$X_1,X_2,\dots,X_n\ i.i.d\sim Exp(\lambda)$，所以$Y=\sum_{i=1}^{n}X_i\sim Ga(n,\lambda)$，相应的密度函数为

$$f(y;n,\lambda)=\frac{\lambda^n}{\Gamma(n)}y^{n-1}\exp(-\lambda y)，\ y>0，$$

于是

$$E(1/Y)=\int_0^{\infty}\frac{\lambda^n}{\Gamma(n)}y^{n-2}e^{-\lambda y}\,dy$$

$$=\frac{\lambda}{n-1}\int_0^{\infty}\frac{\lambda^{n-1}}{\Gamma(n-1)}y^{n-2}e^{-\lambda y}\,dy=\frac{\lambda}{n-1},$$

所以，$E(1/\bar{X})=\frac{n\lambda}{n-1}$，即$1/\bar{X}$不是$\lambda$的无偏估计，但它是$\lambda$的渐进无偏估计，经修偏，$\frac{n-1}{n\bar{X}}$是$\lambda$的无偏估计。

3.1.1 点估计及无偏性

- 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，否则称该参数**不可估**。

- 例3.3 设$X_1, X_2, \dots, X_n$是来自二点分布$b(1,p), 0<p<1$的样本，则$\theta = 1/p$ 不可估。

- **证明：**用反证法，假设$g(X_1,\cdots,X_n)$是$1/p$ 的无偏估计，则有

$$\sum_{X_1,\cdots,X_n} g(X_1,\cdots,X_n)p^{\sum_{i=1}^n X_i}(1-p)^{n-\sum_{i=1}^n X_i} = \frac{1}{p},$$

即 $\sum_{X_1,\cdots,X_n} \mathrm{g}(X_1,\cdots,X_n)p^{\sum_{i=1}^n X_i+1}(1-p)^{n-\sum_{i=1}^n X_i} - 1 = 0$，

上式是p的$n+1$次方程，它最多有$n+1$个实根，而p可在$(0,1)$取无穷多个值，所以不论取什么形式都不能使上述方程在$0<p<1$上成立，这表明$1/p$ 的无偏估计不存在。

3.1.2 有效性

- **定义3.3** 设$\hat{\theta}_1,\hat{\theta}_2$是$\theta$的两个无偏估计，如果对任意的$\theta\in\Theta$，有

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

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。

3.1.2 有效性

- 例3.4 设$X_1, X_2, \ldots, X_n$是取自某总体的样本，记总体均值为μ，总体方差为σ^2，则$\hat{\mu}_1 = X_1$,$\hat{\mu}_2 = \bar{X}$,都是μ的无偏估计，但

$$\mathrm{Var}(\hat{\mu}_1) = \sigma^2, \mathrm{Var}(\hat{\mu}_2) = \sigma^2/n$$

显然，只要$n > 1$，$\hat{\mu}_2$比$\hat{\mu}_1$有效。这表明用全部数据的平均估计总体均值要比只使用部分数据更有效。

3.1.2 有效性

- **例3.5** 设$X_1,\cdots,X_n$是来自均匀总体$U(0,\theta)$的样本，我们对θ感兴趣。

（1）一种估计：人们常用最大观测值$X_{(n)}$来估计θ，由于

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

所以$X_{(n)}$不是θ的无偏估计，而是θ的渐近无偏估计。

经过修偏后可以得到θ的一个无偏估计：

$$\hat{\theta}_1=\frac{n+1}{n}X_{(n)}.$$

（2）另一种估计：由于总体均值为$\theta/2$，通过利用样本均值估计总体均值的思路，我们可以得到θ的另一个无偏估计：

$$\hat{\theta}_2=2\bar{X}.$$

3.1.2 有效性

$$Var(\hat{\theta}_1) = \left(\frac{n+1}{n}\right)^2 Var(X_{(n)}) = \left(\frac{n+1}{n}\right)^2 \frac{n}{(n+1)^2(n+2)}\theta^2$$
$$= \frac{\theta^2}{n(n+2)}.$$

$$Var(\hat{\theta}_2) = 4Var(\bar{X}) = \frac{4}{n}Var(X) = \frac{4}{n}\cdot\frac{\theta^2}{12} = \frac{\theta^2}{3n}$$

由此，当$n > 1$时，$\hat{\theta}_1$比$\hat{\theta}_2$有效。

3.1.2 有效性

- **补3** 设X_1,X_2,X_3服从均匀分布$U(0,\theta)$，试证$4X_{(1)}$及$\frac{4}{3}X_{(3)}$都是θ的无偏估计量，哪个更有效?
- **证：** 由$X\sim U(0,\theta)$可知$X_{(1)}$、$X_{(3)}$的密度函数分别为

$$f_1(x)=3\cdot\left(\frac{\theta-x}{\theta}\right)^2\cdot\frac{1}{\theta}=\frac{3}{\theta^3}(\theta-x)^2,\quad 0<x<\theta,$$

$$f_3(x)=3\cdot\left(\frac{x}{\theta}\right)^2\cdot\frac{1}{\theta}=\frac{3}{\theta^3}x^2,\quad 0<x<\theta,$$

从而

$$E(X_{(1)})=\frac{3}{\theta^3}\int_0^\theta x(\theta-x)^2\,dx=\frac{\theta}{4},$$

$$E(X_{(3)})=\frac{3}{\theta^3}\int_0^\theta x^3\,dx=\frac{3}{4}\theta,$$

3.1.2 有效性

故，由$E(4X_{(1)})=\theta$，$E(\frac{4}{3}X_{(3)})=\theta$知两者均为$\theta$的无偏估计。

又可算得

$$E(X_{(1)}^2)=\frac{3}{\theta^3}\int_0^\theta x^2(\theta-x)^2\,dx=\frac{1}{10}\theta^2,$$

$$E(X_{(3)}^2)=\frac{3}{\theta^3}\int_0^\theta x^4\,dx=\frac{3}{5}\theta^2,$$

从而$Var(X_{(1)})=\frac{1}{10}\theta^2-\frac{1}{16}\theta^2=\frac{3}{80}\theta^2$，$Var(X_{(3)})=\frac{3}{5}\theta^2-\frac{9}{16}\theta^2=\frac{3}{80}\theta^2$，

因此$Var(4X_{(1)})=\frac{3}{5}\theta^2$，$Var\left(\frac{4}{3}X_{(3)}\right)=\frac{\theta^2}{15}$.

故$Var\left(\frac{4}{3}X_{(3)}\right)<Var(4X_{(1)})$，即$\frac{4}{3}X_{(3)}$更有效。

§ 3.2 矩估计及相合性

3.2.1 替换原理和矩法估计

一、矩法估计

替换原理是指用样本矩及其函数去替换相应的总体矩及其函数，譬如：

- 用样本均值估计总体均值$E(X)$，即$\hat{E}(X)=\bar{X}$；
- 用样本方差估计总体方差Var(X)，即$\widehat{V}ar(X)=S^2$；
- 用样本的p分位数估计总体的p分位数；
- 用样本中位数估计总体中位数。

3.2.1 替换原理和矩法估计

- 例3.6 对某型号的20辆汽车记录其每5L汽油的行驶里程(km)，观测数据如下：

29.8	27.6	28.3	27.9	30.1	28.7	29.9	28.0
27.9	28.7	28.4	27.2	29.5	28.5	28.0	30.0
29.1	29.8	29.6	26.9				

经计算有

$$\bar{x} = 28.695, \quad s^2 = 0.9185, \quad m_{0.5} = 28.6$$

由此给出总体均值、方差和中位数的估计分别为: 28.695，0.9185 和 28.6。

3.2.2 概率函数已知时未知参数的矩估计

- 矩估计的实质是用经验分布函数去替换总体分布，其理论基础是格里纹科定理。

- 设总体具有已知的概率函数 $f(x;\theta_1,\dots,\theta_k)$，$(\theta_1,\dots,\theta_k)\in\Theta$是未知参数或参数向量，$X_1,X_2,\dots,X_n$是样本，假定总体的$k$阶原点矩$\mu_k$存在，若$\theta_1,\dots,\theta_k$能够表示成$\mu_1,\dots,\mu_k$的函数$\theta_j=\theta_j(\mu_1,\dots,\mu_k)$，则可给出诸$\theta_j$的矩估计为

$$\hat{\theta}_j=\theta_j(A_1,\dots,A_k),\quad j=1,\dots,k,$$

其中 $A_j=\frac{1}{n}\sum_{i=1}^{n}X_i^j$.

3.2.2 概率函数已知时未知参数的矩估计

- 补4 已知正态总体$N(\mu,\sigma^2)$，求μ和σ^2的矩估计。
- **解法一：** 因为μ和σ^2分别为总体均值和方差，根据替换原理，μ和σ^2的矩估计分别为

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

解法二： $\begin{cases} E(X)=\mu, \\ E(X^2)=Var(X)+[E(X)]^2=\sigma^2+\mu^2. \end{cases}$

于是可得$\begin{cases} \mu=E(X), \\ \sigma^2=E(X^2)-\mu^2. \end{cases}$

由此μ和σ^2的矩估计为

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

3.2.2 概率函数已知时未知参数的矩估计

- 例3.7 设总体服从指数分布，由于 $EX=\frac{1}{\lambda}$，即 $\lambda=1/EX$，故λ的矩估计为

$$\hat{\lambda}=1/\bar{X}.$$

另外，由于$Var(X)=1/\lambda^2$，其反函数为 $\lambda=1/\sqrt{Var(X)}$ 。因此，从替换原理来看，λ 的矩法估计也可取为

$$\hat{\lambda}_1=1/S,$$

S为样本标准差。

这说明矩估计可能是不唯一的，这是矩估计的一个缺点，此时通常应该尽量采用低阶矩给出未知参数的估计。

3.2.2 概率函数已知时未知参数的矩估计

- **例3.8** 设$X_1, X_2, \ldots, X_n$是来自(a,b)上的均匀分布$U(a,b)$的样本，a与b均是未知参数，这里$k=2$，由于

$$E(X)=\frac{a+b}{2}, \qquad \mathrm{Var}(X)=\frac{(b-a)^2}{12},$$

不难推出

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

由此即可得到a,b的矩估计:

$$\hat{a}=\bar{X}-\sqrt{3}S, \qquad \hat{b}=\bar{X}+\sqrt{3}S.$$

3.2.2 概率函数已知时未知参数的矩估计

- 补5 设总体X的概率密度函数为

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

求未知参数θ矩估计量。

- **解：** $E(X)=\int_0^1 x\cdot\theta x^{\theta-1}\,dx=\frac{\theta}{\theta+1}$，

令$E(X)=\bar{X}$，则$\hat{\theta}=\frac{\bar{X}}{1-\bar{X}}$，其中$\bar{X}=\frac{1}{n}\sum_{i=1}^n x_i$，

则$\hat{\theta}$即为参数θ的矩估计。

3.2.3 相合性

- 我们知道，点估计是一个统计量，因此它是一个随机变量，在样本量一定的条件下，我们不可能要求它完全等同于参数的真实取值。
- 但如果我们有足够的观测值，根据格里纹科定理，随着样本量的不断增大，经验分布函数逼近真实分布函数，因此完全可以要求估计量随着样本量的不断增大而逼近参数真值，这就是**相合性**，严格定义如下：

3.2.3 相合性

- **定义3.4** 设$\theta\in\Theta$为未知参数，$\hat{\theta}_n=\hat{\theta}_n(X_1,\ldots,X_n)$是$\theta$的一个估计量，$n$是样本容量，若对任何一个$\varepsilon>0$，有

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

则称$\hat{\theta}_n$为θ参数的**相合估计**。

3.2.3 相合性

- 相合性被认为是对估计的一个最基本要求, 如果一个估计量, 在样本量不断增大时，它都不能把被估参数估计到任意指定的精度, 那么这个估计是很值得怀疑的。 通常, 不满足相合性要求的估计一般不予考虑。
- 若把依赖于样本量n的估计量$\hat{\theta}_n$看作一个随机变量序列, 相合性就是$\hat{\theta}_n$依概率收敛于θ, 所以证明估计的相合性可应用依概率收敛的性质及各种大数定律。

3.2.3 相合性

在判断估计的相合性时下述两个定理是很有用的。

- **定理3.1** 设$\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$是$\theta$的一个估计量，若

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

  则$\hat{\theta}_n$是θ的相合估计。

- **定理3.2** 若$\hat{\theta}_{n1}, \cdots, \hat{\theta}_{nk}$分别是$\theta_1, \cdots, \theta_k$的相合估计，$\eta = g(\theta_1, \cdots, \theta_k)$是$\theta_1, \cdots, \theta_k$的连续函数，则

  $\hat{\eta}_n = g(\hat{\theta}_{n1}, \cdots, \hat{\theta}_{nk})$是$\eta$的相合估计。

3.2.3 相合性

- 例3.9 设$X_1, X_2, \dots, X_n$是来自均匀总体$U(0,\theta)$的样本，证明$X_{(n)}$是θ的相合估计。
- **证明：** 由次序统计量的分布，我们知道$X_{(n)}$的分布密度函数为$f(y) = ny^{n-1}/\theta^n,\ 0 < y < \theta$，故有

$$E(\hat{\theta}) = \int_0^\theta ny^n \mathrm{d}y/\theta^n = \frac{n}{n+1}\theta \to \theta\ (n \to \infty)$$

$$E(\hat{\theta}^2) = \int_0^\theta ny^{n+1} \mathrm{d}y/\theta^n = \frac{n}{n+2}\theta^2$$

$$Var(\hat{\theta}) = \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2 = \frac{n}{(n+1)^2(n+2)}\theta^2 \to 0$$

由定理3.1可知，$X_{(n)}$是θ的相合估计。

3.2.3 相合性

- 补6 设$X_1, X_2, \ldots, X_n$独立同分布，$E(X_1) = \mu$， $Var(X_1) < +\infty$，证明：$\hat{\mu} = \frac{2}{n(n+1)}\sum_{i=1}^{n} iX_i$是$\mu$的相合估计。
- **证明：** 由于

$$E(\hat{\mu}) = \frac{2\mu}{n(n+1)} \cdot \sum_{i=1}^{n} i = \frac{2\mu}{n(n+1)} \cdot \frac{n(n+1)}{2} = \mu,$$

$$Var(\hat{\mu}) = \frac{4Var(X_1)}{n^2(n+1)^2} \cdot \sum_{i=1}^{n} i^2$$

$$= \frac{4Var(X_1)}{n^2(n+1)^2} \cdot \frac{n(n+1)(2n+1)}{6} \to 0(n \to +\infty),$$

这就证明了$\hat{\mu} = \frac{2}{n(n+1)}\sum_{i=1}^{n} iX_i$是$\mu$的相合估计。

3.2.3 相合性

由大数定律及定理3.2，我们可以看到：

矩估计一般都具有相合性。比如：

- 样本均值是总体均值的相合估计；
- 样本标准差是总体标准差的相合估计；
- 样本变异系数是总体变异系数的相合估计。

（例如参考教材例6.2.6）

§3.3 极大似然估计

- **Fisher的极大似然思想：** 一个随机试验有很多可能结果，如果在一次试验中，某结果发生了，则认为该结果（事件）发生的可能性最大。

- 补7 一老战士与一新同学一同进行射击训练，每人打了一枪，结果有一枪中靶。试问这一枪是谁打中的？
- 分析：按照Fisher的极大似然思想，应该认为是老战士打中的较合理。

§3.3 极大似然估计

- **补8** 一袋中有红、白两颜色的球若干，只知道两种球的比例为4：1，但不知道哪种颜色的球占多。现从中任取一球，结果为白色。问袋中哪种颜色的球较多？
- **分析：** 按照Fisher的极大似然思想，应该认为袋中白球较多。

3.3.1 极大似然估计

- **例3.10**设有外形完全相同的两个箱子，甲箱中有99个白球和1个黑球，乙箱中有99个黑球和1个白球，今随机地抽取一箱，并从中随机抽取一球，结果取得白球，问这球是从哪一个箱子中取出？
- **分析：**不管是哪一个箱子，从箱子中任取一球都有两个可能的结果：A 表示“取出白球”，B 表示“取出黑球”。如果我们取出的是甲箱，则A发生的概率为0.99，而如果取出的是乙箱，则A发生的概率为0.01。

3.3.1 极大似然估计

现在一次试验中结果A发生了，人们的第一印象就是："此白球（A）最像从甲箱中取出的"，或者说，应该认为试验条件对结果A出现有利，从而可以推断这球是从甲箱中取出的。这个推断 很符合人们的经验事实，这里"最像"就是"极大似然"之意。这种想法常称为"极大似然原理"。

3.3.1极大似然估计

- 补9 考察以下例子：

假设在一个罐中放着许多白球和黑球，并假定已经知道两种球的数目之比是1：3，但不知道哪种颜色的球多。如果用返回抽样方法从罐中任取n个球，则其中黑色的个数为x的概率为：

$$P(x,p)=\binom{n}{x}p^{x}q^{(n-x)},$$

$$\text{其中}q=1-p,\text{由假设知，}\ p=\frac{1}{4}\text{或}\frac{3}{4}$$

若取$n=3$，如何通过x来估计p值？

3.3.1 极大似然估计

❖分析：

先计算抽样的可能结果x在这两种p值之下的概率：

x	0	1	2	3
$P(x,\frac{3}{4})$	$\frac{1}{64}$	$\frac{9}{64}$	$\frac{27}{64}$	$\frac{27}{64}$
$P(x,\frac{1}{4})$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$

3.3.1 极大似然估计

从上表看到：

$x=0, P\left(0,\frac{1}{4}\right)=\frac{27}{64}>P\left(0,\frac{3}{4}\right)=\frac{1}{64}$, 取$p=\frac{1}{4}$更合理；

$x=1$类似；

$x=2, P\left(2,\frac{1}{4}\right)=\frac{9}{64}<P\left(2,\frac{3}{4}\right)=\frac{27}{64}$, 取$p=\frac{3}{4}$更合理；

$x=3$类似；

于是有：$\hat{p}(x)=\begin{cases}\frac{1}{4} & x=0,1\\ \frac{3}{4} & x=2,3\end{cases}$

3.3.1 极大似然估计

我们在选取上述估计时遵循了如下原理：认为概率最大的事件最有可能发生，即我们选取的估计$\hat{p}(x)$满足如下不等式：

$$P\{X=x|\hat{p}(x)\}\geq P\{X=x|p'\},\forall p'.$$

这就是我们要讲的极大似然原理的基本思想。

3.3.1 极大似然估计

- 例3.11 设产品分为合格品与不合格品两类，我们用一个随机变量X来表示某个产品经检查后的不合格品数，则$X=0$表示合格品，$X=1$表示不合格品，则X服从二点分布$b(1,p)$，其中p是未知的不合格品率。现抽取n个产品看其是否合格，得到样本$X_1,\ldots,X_n$。这批观测值发生的概率为

$$P(X_1=x_1,\ldots,X_n=x_n;p)=\prod_{i=1}^{n}p^{x_i}(1-p)^{1-x_i}$$
$$=p^{\sum_{i=1}^{n}x_i}(1-p)^{n-\sum_{i=1}^{n}x_i},$$

由于p是未知的，根据极大似然原理，我们应选择p使得上式表示的概率尽可能大。

3.3.1 极大似然估计

将上式看作未知参数 p 的函数，用 $L(p)$ 表示，称作**似然函数**，亦即

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

要求似然函数的最大值点不是难事，将上式两端取对数并关于p求导令其为0，即得如下方程，又称**似然方程**：

$$\frac{\partial \ln L(p)}{\partial p}=\frac{\sum_{i=1}^{n}x_i}{p}-\frac{n-\sum_{i=1}^{n}x_i}{1-p}=0.$$

解之即得p的**极大似然估计**，为

$$\hat{p}=\hat{p}(X_1,\ldots,X_n)=\sum_{i=1}^{n}X_i/n=\bar{X}.$$

3.3.1 极大似然估计

- **定义3.5** 设总体的概率函数为$f(x,\theta)$，Θ是参数θ可能取值的参数空间，$X_1,X_2,\dots,X_n$是样本，将样本的联合概率函数看成θ的函数用$L(\theta;x_1,x_2,\dots,x_n)$表示，简记为$L(\theta)$，

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

$$=f(x_1;\theta)\cdot f(x_2;\theta)\cdot\cdots\cdot f(x_n;\theta)$$

称为样本的似然函数。

3.3.1 极大似然估计

- 如果某统计量$\hat{\theta}=\hat{\theta}(X_1,\dots,X_n)$满足

$$L(\hat{\theta})=\max_{\theta\in\Theta} L(\theta)$$

则称$\hat{\theta}$是θ的**极(最)大似然估计**，简记为 MLE（Maximum Likelihood Estimate）。

- 人们通常更习惯于由对数似然函数$\ln L(\theta)$出发寻找θ的极大似然估计。

当$L(\theta)$是可微函数时，求导是求极大似然估计最常用的方法，对$\ln L(\theta)$求导更加简单些。

3.3.1 极大似然估计

- 例3.12 设一个试验有三种可能结果，其发生概率分别为 $p_1=\theta^2, p_2=2\theta(1-\theta), p_3=(1-\theta)^2$现做了$n$ 次试验，观测到三种结果发生的次数分别为$n_1,n_2,n_3(n_1+n_2+n_3=n)$，则似然函数为

$$L(\theta)=(\theta^2)^{n_1}[2\theta(1-\theta)]^{n_2}[(1-\theta)^2]^{n_3}$$
$$=2^{n_2}\theta^{2n_1+n_2}(1-\theta)^{2n_3+n_2}$$

其对数似然函数为

$$\ln L(\theta)=(2n_1+n_2)\ln\theta+(2n_3+n_2)\ln(1-\theta)+n_2\ln 2$$

3.3.1 极大似然估计

将之关于θ求导，并令其为0得到似然方程

$$\frac{2n_1+n_2}{\theta}-\frac{2n_3+n_2}{1-\theta}=0$$

解之，得

$$\hat{\theta}=\frac{2n_1+n_2}{2(n_1+n_2+n_3)}=\frac{2n_1+n_2}{2n}$$

由于

$$\frac{\partial^2 \ln L(\theta)}{\partial\theta^2}=-\frac{2n_1+n_2}{\theta^2}-\frac{2n_3+n_2}{(1-\theta)^2}<0$$

所以$\hat{\theta}$是极大值点。故$\hat{\theta}$为θ的极大似然估计。

3.3.1 极大似然估计

- 例3.13 对正态总体$N(\mu,\sigma^2)$，$\theta=(\mu,\sigma^2)$是二维参数，设有样本$X_1,X_2,\dots,X_n$，则似然函数及其对数分别为

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

$$=(2\pi\sigma^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

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

3.3.1 极大似然估计

将$\ln L(\mu,\sigma^2)$分别关于两个分量求偏导并令其为0，即得到似然方程组

$$\frac{\partial \ln L(\mu,\sigma^2)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)=0$$

$$\frac{\partial \ln L(\mu,\sigma^2)}{\partial \sigma^2} = \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2-\frac{n}{2\sigma^2}=0$$

3.3.1 极大似然估计

解此方程组，由第一个方程可得μ的极大似然估计为

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

将之代入第二个方程，得出σ^2的极大似然估计

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

利用二阶导函数矩阵的非正定性可以说明上述估计使得似然函数取极大值。

3.3.1 极大似然估计

- 虽然求导函数是求极大似然估计最常用的方法，但并不是在所有场合求导都是有效的。
- 例3.14 设 $X_1, X_2, \dots, X_n$是来自均匀总体$U(0,\theta)$的样本，试求θ的极大似然估计。

3.3.1 极大似然估计

- **解:** 似然函数

$$L(\theta)=\frac{1}{\theta^n}\prod_{i=1}^{n} I_{\{0<x_i\le\theta\}}=\frac{1}{\theta^n}\prod_{i=1}^{n} I_{\{x_{(n)}\le\theta\}}$$

要使$L(\theta)$达到最大，首先一点是示性函数取值应该为1，其次是$1/\theta^n$尽可能大。由于$1/\theta^n$是θ的单调减函数，所以θ的取值应尽可能小，但示性函数为1决定了θ不能小于$X_{(n)}$，由此给出θ的极大似然估计：$\hat{\theta}=X_{(n)}$。

3.3.1 极大似然估计

- 极大似然估计有一个简单而有用的性质：如果$\hat{\theta}$是θ的极大似然估计，则对任一函数$g(\theta)$，其极大似然估计为$g(\hat{\theta})$。
- 该性质称为极大似然估计的**不变性**，从而使一些复杂结构的参数的极大似然估计的获得变得容易了。

3.3.1 极大似然估计

- 例3.15 设$X_1, X_2, \ldots, X_n$是来自正态总体$N(\mu, \sigma^2)$的样本，则μ和σ^2的极大似然估计为$\hat{\mu} = \bar{X}$，$\hat{\sigma}^2 = S_n^2$，于是由不变性可得如下参数的极大似然估计，它们是:

- 标准差σ的MLE是$\hat{\sigma} = S_n$；
- 概率$P(X < 3) = \Phi\left(\frac{3-\mu}{\sigma}\right)$的MLE是$\Phi\left(\frac{3-\bar{X}}{S_n}\right)$；
- 总体0.90分位数$X_{0.90} = \mu + \sigma u_{0.90}$的MLE是$\bar{X} + S_n u_{0.90}$，其中$u_{0.90}$为标准正态分布的0.90分位数。

3.3.1 极大似然估计

- **补10** 设$X_1, X_2, \dots, X_n$是来自密度函数为$f(x;\theta) = e^{-(x-\theta)}$，$x > \theta$的样本，求$\theta$的极大似然估计$\hat{\theta}$，它是否是相合估计？是否是无偏估计？
- **解：** 参数θ的似然函数为

$$L(\theta) = \prod_{i=1}^{n} \{e^{-(x_i-\theta)} I_{\{x_i>\theta\}}\}$$

$$= \exp\{-\sum_{i=1}^{n} x_i + n\theta\} I_{\{x_{(1)}>\theta\}}.$$

显然$L(\theta)$在示性函数为1的条件下是θ的严格增函数，因此θ的最大似然估计为$\hat{\theta} = X_{(1)}$。

3.3.1 极大似然估计

又$X_{(1)}$的密度函数为$f(x)=ne^{-n(x-\theta)}$，$x>\theta$，故

$$E(\hat{\theta})=\int_{\theta}^{+\infty} xne^{-n(x-\theta)}\,dx$$

$$=\int_{0}^{+\infty}(t+\theta)ne^{-nt}\,dt=\frac{1}{n}+\theta,$$

故$\hat{\theta}$不是θ的无偏估计，但是θ的渐进无偏估计。

由于 $E(\hat{\theta})\to\theta(n\to+\infty)$且

3.3.1 极大似然估计

$$E(\hat{\theta}^2) = \int_{\theta}^{+\infty} x^2 n e^{-n(x-\theta)}\, dx$$

$$= \int_{0}^{+\infty} (t^2 + 2\theta t + \theta^2) n e^{-nt}\, dt$$

$$= \frac{2}{n^2} + \frac{2}{n}\theta + \theta^2,$$

$$Var(\hat{\theta}) = \frac{2}{n^2} + \frac{2\theta}{n} + \theta^2 - \left(\frac{1}{n} + \theta\right)^2 = \frac{1}{n^2} \to 0,$$

这说明$\hat{\theta}$是θ的相合估计。

3.3.3 渐近正态性

- 极大似然估计有一个良好的性质：它通常具有渐近正态性。
- **定义3.6** 参数θ的相合估计$\hat{\theta}_n$称为渐近正态的，若存在趋于0的非负常数序列$\sigma_n(\theta)$, 使得 $\frac{\hat{\theta}_n-\theta}{\sigma_n(\theta)}$ 依分布收敛于标准正态分布，称$\hat{\theta}_n$ 服从**渐近正态分布** $N\left(\theta,\ {\sigma_n}^2(\theta)\right)$, 记为$\hat{\theta}_n \sim AN\left(\theta,\ {\sigma_n}^2(\theta)\right)$。称${\sigma_n}^2(\theta)$为$\hat{\theta}_n$的**渐近方差**。
- 参考教材例6.3.8：泊松分布参数估计。

§3.4 最小方差无偏估计

3.4.1 均方误差

- 无偏估计不一定比有偏估计更优。
- 评价一个点估计的好坏一般可以用：点估计值$\hat{\theta}$与参数真值θ的距离平方的期望，这就是下式给出的**均方误差**

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

均方误差是评价点估计的最一般的标准。我们希望估计的均方误差越小越好。

3.4.1 均方误差

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

(1) 若$\hat{\theta}$是θ的无偏估计，则 $MSE(\hat{\theta}) = \mathrm{Var}(\hat{\theta})$，这说明用方差考察无偏估计有效性是合理的。

(2) 当 $\hat{\theta}$不是θ 的无偏估计时，就要看其均方误差 $MSE(\hat{\theta})$。

下面的例子说明：在均方误差的含义下有些有偏估计优于无偏估计。

3.4.1 均方误差

❖ 例3.16 对均匀总体$U(0,\theta)$，由θ的极大似然估计得到的无偏估计是$\hat{\theta}=(n+1)X_{(n)}/n$，它的均方误差

$$MSE(\hat{\theta})=\mathrm{Var}(\hat{\theta})=\frac{\theta^2}{n(n+2)}$$

现我们考虑的θ形如$\hat{\theta}_\alpha=\alpha\cdot X_{(n)}$的估计，其均方误差为

$$MSE(\hat{\theta}_\alpha)=\alpha^2\frac{n}{(n+1)^2(n+2)}\theta^2+\left(\frac{n\cdot\alpha}{n+1}-1\right)^2\theta^2$$

用求导的方法不难求出当$\alpha_0=(n+2)/(n+1)$时上述均方误差达到最小，且其均方误差

$$MSE(\hat{\theta}_0)=\frac{\theta^2}{(n+1)^2}<\frac{\theta^2}{n(n+2)}=MSE(\hat{\theta})$$

所以在均方误差的标准下，有偏估计$\hat{\theta}_0$优于无偏估计$\hat{\theta}$。

3.4.1 均方误差

- **补11** 设$X_1,\dots,X_n$抽自均值为μ的总体，考虑μ的如下两个估计

$$\hat{\mu}=\bar{X},$$

$$\hat{\mu}_{(-i)}=\sum_{j\neq i}X_j/(n-1),$$

这里$\hat{\mu}_{(-i)}$表示去掉第i个样本X_i后，对其余n-1个样本所求的样本均值。显然，$\hat{\mu}$和$\hat{\mu}_{(-i)}$都是μ的无偏估计，但是因为$X_1,\dots,X_n$都是独立同分布的，且$Var(X_i)=\sigma^2$，于是

3.4.1 均方误差

$$Var(\hat{\mu}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{\sigma^2}{n},$$

$$Var(\hat{\mu}_{(-i)}) = \frac{\sigma^2}{n-1}.$$

可见，$\hat{\mu} = \bar{X}$比$\hat{\mu}_{(-i)}$有较小的方差，因而$\bar{X}$优于$\hat{\mu}_{(-i)}$。这表明，当我们用样本均值去估计总体均值时，使用全体样本总比不使用全体样本要好。

3.4.1 均方误差

- 补12 *设总体概率密度函数是$f(x;\theta)$, $X_1, X_2, \ldots, X_n$是其样本，$T=T(X_1,\ldots,X_n)$是θ的充分统计量，则对$g(\theta)$的任一估计$\hat{g}$，令$\tilde{g}=E(\hat{g}|T)$，证明：$MSE(\tilde{g})\le MSE(\hat{g})$.这说明，在均方误差准则下，人们只需要考虑基于充分统计量的估计。

- **证明：** 我们将均方误差作如下分解

$$
\begin{aligned}
MSE(\hat{g}) &= E(\hat{g}-\theta)^2 = E(\hat{g}-\tilde{g}+\tilde{g}-\theta)^2 \\
&= E(\hat{g}-\tilde{g})^2 + MSE(\tilde{g}) + 2E[(\hat{g}-\tilde{g})(\tilde{g}-\theta)].
\end{aligned}
$$

注意到，$\tilde{g}=E(\hat{g}|T)$，这说明

$$
\begin{aligned}
E[(\hat{g}-\tilde{g})|T] &= E(\hat{g}|T) - E[E(\hat{g}|T)|T] \\
&= E(\hat{g}|T) - E(\hat{g}|T) = 0,
\end{aligned}
$$

3.4.1 均方误差

于是

$$\mathrm{E}[(\hat{\mathrm{g}}-\tilde{g})(\tilde{\mathrm{g}}-\theta)] = E\{E[(\hat{g}-\tilde{g})(\tilde{g}-\theta)|T]\}$$
$$= E\{(\tilde{g}-\theta)E[(\hat{g}-\tilde{g})|T]\} = 0,$$

因而$MSE(\hat{g}) = E(\hat{g}-\tilde{g})^2 + MSE(\tilde{g}) \geq MSE(\tilde{g})$.

3.4.2 一致最小方差无偏估计*

- **定义3.7** 对参数估计问题，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Θ上都有

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

则称$\hat{\theta}$是θ的一致最小方差无偏估计，简记为UMVUE。如果UMVUE存在，则它一定是充分统计量的函数。

3.4.2 一致最小方差无偏估计*

关于UMVUE，有如下一个判断准则。

- **定理3.3** 设$\boldsymbol{X}=(X_1,X_2,\ldots,X_n)$是来自某总体的一个样本，$\hat{\theta}=\hat{\theta}(\boldsymbol{X})$是$\theta$的一个无偏估计，$\mathrm{Var}(\hat{\theta})<\infty$。则$\hat{\theta}$是$\theta$的UMVUE的充要条件是，对任意一个满足$E\left(\varphi(\boldsymbol{X})\right)=0$和$\mathrm{Var}\left(\varphi(\boldsymbol{X})\right)<\infty$的$\varphi(\boldsymbol{X})$，都有

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

3.4.2 一致最小方差无偏估计*

- 例3.17 设$X_1,\dots,X_n$是来自指数分布$Exp(1/\theta)$的样本，则$T=X_1+\cdots+X_n$是θ的充分统计量，而$\bar{X}=T/n$是θ的无偏估计。设$\varphi=\varphi(X_1,\dots,X_n)$是0的任一无偏估计，则

$$\int_0^\infty \cdots \int_0^\infty \varphi(x_1,\dots,x_n)\cdot e^{-(x_1+\dots+x_n)/\theta}\,dx_1\cdots dx_n=0$$

两端对θ求导得

$$\int_0^\infty \cdots \int_0^\infty \frac{n\bar{x}}{\theta^2}\varphi(x_1,\dots,x_n)\cdot e^{-(x_1+\dots+x_n)/\theta}\,dx_1\cdots dx_n=0$$

这说明$E(\bar{x}\cdot\varphi)=0$，从而

$$\mathrm{Cov}(\bar{x},\varphi)=E(\bar{x}\cdot\varphi)-\mathrm{E}(\bar{x})\cdot E(\varphi)=0,$$

由定理3.3，$\bar{X}$是θ的UMVUE。

3.4.2 一致最小方差无偏估计*

- **补13** 设T_1, T_2分别是θ_1, θ_2的UMVUE，证明：对任意的（非零）常数a, b, $aT_1 + bT_2$是$a\theta_1 + b\theta_2$的UMVUE。

- **证明：** 由于T_1, T_2分别是θ_1, θ_2的UMVUE，故$E(T_i) = \theta_i$，且对任意一个$\phi(x)$，满足$E\phi = 0$，由判断准则知$Cov(T_i, \phi) = 0, i = 1,2$，于是
$$E(aT_1 + bT_2) = a\theta_1 + b\theta_2,$$
$$Cov(aT_1 + bT_2, \phi) = aCov(T_1, \phi) + bCov(T_2, \phi) = 0,$$
因此$aT_1 + bT_2$是$a\theta_1 + b\theta_2$的UMVUE。

3.4.3 充分性原则*

以下定理说明：

较好的无偏估计都是充分统计量的函数。

- 定理3.4 设总体概率函数$p(x,\theta)$， $X_1,\dots,X_n$是其样本，$T=T(X_1,\dots,X_n)$是θ的充分统计量，则对θ的任一无偏估计$\hat{\theta}=\hat{\theta}(X_1,\dots,X_n)$，令$\tilde{\theta}=E\left(\hat{\theta}\middle|T\right)$，则$\tilde{\theta}$也是$\theta$的无偏估计，且

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

3.4.3 充分性原则*

- **定理3.4说明：** 如果无偏估计不是充分统计量的函数，则将之对充分统计量求条件期望可以得到一个新的无偏估计，该估计的方差比原来的估计的方差要小，从而降低了无偏估计的方差。换言之，考虑θ的估计问题只需要在基于充分统计量的函数中进行即可，该说法对所有的统计推断问题都是正确的，这便是所谓的**充分性原则**。

3.4.3 充分性原则*

- 例3.18 设$X_1,\ldots,X_n$是来自$b(1,p)$的样本，则 $T=n\bar{X}$是p的充分统计量。为估计$\theta=p^2$，可令

$$\hat{\theta}_1=\begin{cases}1, & x_1=1,x_2=1\\ 0, & \text{其它}\end{cases}$$

由于$E(\hat{\theta}_1)=P(X_1=1,X_2=1)=p\cdot p=\theta$，所以$\hat{\theta}_1$是$\theta$的无偏估计。这个只使用了两个观测值的估计并不好.下面我们用定理3.4对之加以改进：求$\hat{\theta}_1$关于充分统计量$T=\sum_{i=1}^n X_i$的条件期望，得

$$\hat{\theta}=E(\hat{\theta}_1|T=t)=\frac{\binom{n-2}{t-2}}{\binom{n}{t}}=\frac{t(t-1)}{n(n-1)}，\text{其中}t=\sum_{i=1}^n x_i\,.$$

可以验证$\hat{\theta}$是θ的无偏估计，且$\mathrm{Var}(\hat{\theta})\le\mathrm{Var}(\hat{\theta}_1)$

3.4.4 Cramer-Rao不等式*

- **定义3.8** 设总体的概率函数$f(x;\theta)$，$\theta\in\Theta$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 $S=\{x: f(x;\theta)>0\}$与θ无关；
  (3) 导数 $\frac{\partial}{\partial\theta}f(x;\theta)$对一切$\theta\in\Theta$都存在；
  (4) 对$f(x;\theta)$，积分与微分运算可交换次序；
  (5) 期望$E\left[\frac{\partial}{\partial\theta}\ln f(x;\theta)\right]^2$存在；

则称$I(\theta)=E\left[\frac{\partial}{\partial\theta}\ln f(x;\theta)\right]^2$为总体分布的**费希尔(Fisher)信息量。**

3.4.4 Cramer-Rao不等式*

- 费希尔信息量是数理统计学中一个基本概念，很多的统计结果都与费希尔信息量有关。如极大似然估计的渐近方差，无偏估计的方差的下界等都与费希尔信息量$I(\theta)$有关。$I(\theta)$的种种性质显示，“$I(\theta)$越大”可被解释为总体分布中包含未知参数θ的信息越多。

3.4.4 Cramer-Rao不等式*

- **例3.19** 设总体为泊松分布$P(\lambda)$分布，则

$$\ln f(x;\lambda) = x\ln\lambda - \lambda - \ln(x!)$$

$$\frac{\partial}{\partial\lambda}\ln f(x;\lambda) = \frac{x}{\lambda} - 1$$

于是

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

3.4.4 Cramer-Rao不等式*

❖ 例3.20 设总体为指数分布，其密度函数为

$$f(x;\theta)=\frac{1}{\theta}\exp\left\{-\frac{x}{\theta}\right\},x>0,\theta>0$$

可以验证定义3.8的条件满足，且

$$\frac{\partial}{\partial\theta}\ln f(x;\theta)=-\frac{1}{\theta}+\frac{x}{\theta^2}=\frac{x-\theta}{\theta^2}$$

于是

$$I(\theta)=E\left(\frac{x-\theta}{\theta^2}\right)^2=\frac{\mathrm{Var}(x)}{\theta^4}=\frac{1}{\theta^2}$$

3.4.4 Cramer-Rao不等式*

- **补14**设总体的概率函数为$f(x;\theta)$的费希尔信息量存在，若二阶导数$\frac{\partial^2}{\partial\theta^2}f(x;\theta)$对一切的$\theta\in\Theta$存在，证明费希尔信息量$I(\theta)=-E(\frac{\partial^2}{\partial\theta^2}\ln f(x;\theta))$.
- **证明：** 记$S_\theta=\frac{\partial\ln f(x;\theta)}{\partial\theta}$，则

$$ES_\theta=\int_{-\infty}^{+\infty}\frac{1}{f(x;\theta)}\cdot\frac{\partial f(x;\theta)}{\partial\theta}\cdot f(x;\theta)\,dx$$

$$=\int_{-\infty}^{+\infty}\frac{\partial f(x;\theta)}{\partial\theta}dx=\frac{\partial}{\partial\theta}\int_{-\infty}^{+\infty}f(x;\theta)\,dx=0,$$

所以，

3.4.4 Cramer-Rao不等式*

$$\frac{\partial ES_\theta}{\theta} = \frac{\partial}{\partial\theta}\int_{-\infty}^{+\infty} S_\theta f(x;\theta)\,dx = \int_{-\infty}^{+\infty} \frac{\partial}{\partial\theta}\left(S_\theta \cdot f(x;\theta)\right)dx$$

$$= \int_{-\infty}^{+\infty} \left(\frac{\partial S_\theta}{\partial\theta}\cdot f(x;\theta) + S_\theta \cdot \frac{\partial f(x;\theta)}{\partial\theta}\right)dx$$

$$= \int_{-\infty}^{+\infty} \frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2}\cdot f(x;\theta)\,dx$$

$$+ \int_{-\infty}^{+\infty} \left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right)^2 \cdot f(x;\theta)\,dx$$

$$= E\left(\frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2}\right) + ES_\theta^2 = E\left(\frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2}\right) + I(\theta).$$

这就证明了$I(\theta) = -E\left(\frac{\partial^2}{\partial\theta^2}\ln f(x;\theta)\right)$.

3.4.4 Cramer-Rao不等式*

❖ **定理3.5（Cramer-Rao不等式）**

设总体分布$f(x;\theta)$满足定义3.8的条件，$X_1,\dots,X_n$是来自该总体的样本，$T=T(X_1,\dots,X_n)$是$\mathrm{g}(\theta)$的任一个无偏估计，$\mathrm{g}'(\theta)=\frac{\partial g(\theta)}{\partial\theta}$存在，且对$\Theta$中一切$\theta$，微分可在积分号下进行，则有

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

称为**克拉默-拉奥（C-R）不等式**；$[\mathrm{g}'(\theta)]^2/(nI(\theta))$称为$\mathrm{g}(\theta)$的无偏估计的方差的C-R下界，简称$\mathrm{g}(\theta)$的**C-R下界**。

3.4.4 Cramer-Rao不等式*

- 特别，对 θ 的无偏估计 $\hat{\theta}$，有 $\mathrm{Var}\left(\hat{\theta}\right) \geq (nI(\theta))^{-1}$；
- 如果等号成立，则称$T = T(X_1, \ldots, X_n)$是g(θ)的有效估计，有效估计一定是UMVUE。

3.4.4 Cramer-Rao不等式*

- 例3.21 设总体分布列为$f(x;\theta)=\theta^x(1-\theta)^{1-x}$，$x=0,1$，它满足定义3.8的所有条件，可以算得该分布的费希尔信息量为$I(\theta)=\frac{1}{\theta(1-\theta)}$，若$X_1,\dots,X_n$是该总体的样本，则$\theta$的C-R下界为$(nI(\theta))^{-1}=\theta(1-\theta)/n$。因为$\bar{X}$是$\theta$的无偏估计，且其方差等于$\theta(1-\theta)/n$，达到C-R 下界，所以$\bar{X}$是$\theta$的有效估计，它也是$\theta$的UMVUE。

3.4.4 Cramer-Rao不等式*

- 例3.22 设总体为指数分布$Exp(1/\theta)$，它满足定义3.8的所有条件，例3.20中已经算出该分布的费希尔信息量为$I(\theta)=\theta^{-2}$，若$X_1,\dots,X_n$是样本，则θ的C-R下界为$(nI(\theta))^{-1}=\theta^2/n$。而$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$是$\theta$的无偏估计，且其方差等于$\theta^2/n$，达到了C-R下界，所以，$\bar{X}$是$\theta$的有效估计，它也是$\theta$的UMVUE。

3.4.4 Cramer-Rao不等式*

- **能达到C-R下界的无偏估计并不多:**
- 例3.23 设总体为$N(0,\sigma^2)$，满足定义3.8的条件，且费希尔信息量为$I(\sigma^2)=\frac{1}{2\sigma^4}$，令$\sigma=g(\sigma^2)=\sqrt{\sigma^2}$，则$\sigma$的C-R下界为$\frac{[g'(\sigma^2)]^2}{nI(\sigma^2)}=\frac{\sigma^2}{2n}$，

  而σ的UMVUE为$\hat{\sigma}=\sqrt{\frac{n}{2}}\cdot\frac{\Gamma(n/2)}{\Gamma((n+1)/2)}\sqrt{\frac{1}{n}\sum_{i=1}^{n}X_i^2}$

  其方差大于C-R下界。这表明所有σ的无偏估计的方差都大于其C-R下界。

3.4.4 Cramer-Rao不等式*

- **补15** 设$X_1,\dots,X_n$是来自$Ga(\alpha,\lambda)$的样本，$\alpha>0$已知，试证明，$\bar{X}/\alpha$是$g(\lambda)=1/\lambda$的有效估计，从而也是UMVUE。
- **证明：** 总体$Ga(\alpha,\lambda)$的密度函数为

$$f(x;\lambda)=\frac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x},\ x>0,$$

于是

$$\ln p(x;\lambda)=\alpha\ln\lambda-\ln\Gamma(\alpha)+(\alpha-1)\ln x-\lambda x,$$

$$\frac{\partial\ln p(x;\lambda)}{\partial\lambda}=\frac{\alpha}{\lambda}-x,\ \frac{\partial^2\ln p(x;\lambda)}{\partial\lambda^2}=-\frac{\alpha}{\lambda^2},$$

3.4.4 Cramer-Rao不等式*

所以λ的费希尔信息量为$I(\lambda)=\frac{\alpha}{\lambda^2}$，这就是说

$g(\lambda)=\frac{1}{\lambda}$的任一无偏估计的C-R下界为

$$\frac{(g'(\lambda))^2}{nI(\lambda)}=\frac{1}{n\alpha\lambda^2}.$$

又$E\left(\frac{\bar{X}}{\alpha}\right)=\frac{1}{\alpha}\cdot\frac{\alpha}{\lambda}=\frac{1}{\lambda},Var\left(\frac{\bar{X}}{\alpha}\right)=\frac{1}{\alpha^2}\cdot\frac{1}{n}\cdot\frac{\alpha}{\lambda^2}=\frac{1}{n\alpha\lambda^2}.$

这就证明了$\bar{X}/\alpha$是$g(\lambda)=1/\lambda$的有效估计，从而也是UMVUE。

§3.5 贝叶斯估计*

3.5.1 统计推断的基础

- 经典学派的观点：统计推断是根据样本信息对总体分布或总体的特征数进行推断，这里用到两种信息：总体信息和样本信息；
- 贝叶斯学派的观点：除了上述两种信息以外，统计推断还应该使用第三种信息：先验信息。

3.5.1 统计推断的基础*

（1）**总体信息:**总体分布或总体所属分布族提供的信息。

（2）**样本信息:**抽取样本所得观测值提供的信息。

（3）**先验信息:**人们在试验之前对要做的问题在经验上和资料上总是有所了解的，这些信息对统计推断是有益的。先验信息即是抽 样（试验）之前有关统计问题的一些信息。一般说来，先验信息来源于经验和历史资料。先验信息在日常生活和工作中是很重要的。

3.5.1 统计推断的基础*

- 基于上述三种信息进行统计推断的统计学称为**贝叶斯统计学**。它与经典统计学的差别就在于是否利用先验信息。贝叶斯统计在重视使用总体信息和样本信息的同时，还注意先验信息的收集、挖掘和加工，使它数量化，形成先验分布，参加到统计推断中来，以提高统计推断的质量。忽视先验信息的利用，有时是一种浪费，有时还会导出不合理的结论。

3.5.1 统计推断的基础*

- **贝叶斯学派的基本观点：任一未知量θ都可看作随机变量，可用一个概率分布去描述，这个分布称为先验分布；**在获得样本之后，总体分布、样本与先验分布通过贝叶斯公式结合起来得到一个关于未知量θ的新分布—**后验分布；**任何关于θ的统计推断都应该基于θ的后验分布进行。

3.5.2 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$F(\boldsymbol{x}|\theta)$，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的先验信息可确定**先验分布$\boldsymbol{\pi}(\boldsymbol{\theta})$**；
- 从贝叶斯观点看，样本$\boldsymbol{X}=(X_1,\ldots,X_n)$的产生分两步进行:首先从先验分布$\pi(\theta)$产生一个样本$\theta_0$，然后从$F(\boldsymbol{x}|\theta_0)$中产生一组样本。这时样本的**联合条件概率函数**为

$$F(\boldsymbol{x}|\theta_0)=f(x_1,\ldots,x_n|\theta_0)=\prod_{i=1}^{n}f(x_i|\theta_0),$$

这个分布综合了总体信息和样本信息；

3.5.2 贝叶斯公式的密度函数形式*

- 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的其它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$\boldsymbol{X}$和参数θ的**联合分布为:**

$$h(x,\theta)=f(x|\theta)\pi(\theta)$$

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；

3.5.2 贝叶斯公式的密度函数形式*

- 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$\boldsymbol{x}=(x_1,\ldots,x_n)$之后，则应依据$h(\boldsymbol{x},\theta)$对$\theta$作出推断。由于

$$h(\boldsymbol{x},\theta)=\pi(\theta|\boldsymbol{x})m(\boldsymbol{x}),$$

其中$m(\boldsymbol{x})$是$\boldsymbol{x}$的边际概率函数

$$m(\boldsymbol{x})=\int_{\Theta}h(\boldsymbol{x},\theta)\,d\theta=\int_{\Theta}f(\boldsymbol{x}|\theta)\pi(\theta)\,d\theta$$

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$\pi(\theta|\boldsymbol{x})$，它的计算公式是

$$\pi(\theta|\boldsymbol{x})=\frac{h(\boldsymbol{x},\theta)}{m(\boldsymbol{x})}=\frac{f(\boldsymbol{x}|\theta)\pi(\theta)}{\int_{\Theta}f(\boldsymbol{x}|\theta)\pi(\theta)\,d\theta}$$

这个条件分布称为θ的**后验分布.**

3.5.2 贝叶斯公式的密度函数形式*

- 后验分布$\pi(\theta|\boldsymbol{x})$集中了总体、样本和先验中有关$\theta$的一切信息。

- 后验分布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。

3.5.3 贝叶斯估计*

基于后验分布$\pi(\theta|\boldsymbol{x})$估计$\theta$有三种常用的方法：

- 使用后验分布的密度函数最大值作为θ的点估计的**最大后验估计**；
- 使用后验分布的中位数作为θ的点估计的**后验中位数估计**；
- 使用后验分布的均值作为θ的点估计的**后验期望估计**。

用得最多的是后验期望估计，它一般也简称为贝叶斯估计，记为$\hat{\theta}_B$。

3.5.3 贝叶斯估计*

- 例3.24 设某事件A在一次试验中发生的概率为θ，为估计θ，对试验进行了n次独立观测，其中事件A发生了X次，显然$X|\theta \sim b(n,\theta)$，即

$$P(X=x|\theta)=\binom{n}{x}\theta^x(1-\theta)^{n-x},\quad x=0,1,\cdots,n$$

假若我们在试验前对事件A没有什么了解，从而对其发生的概率也没有任何信息。在这种场合，贝叶斯本人建议采用“同等无知”的原则使用区间(0,1)上的均匀分布$U(0,1)$作为θ的先验分布，因为它取(0,1)上的每一点的机会均等。贝叶斯的这个建议被后人称为贝叶斯假设。

3.5.3 贝叶斯估计*

由此即可利用贝叶斯公式求出θ的后验分布。具体如下：先写出X和θ的联合分布

$$h(x,\theta)=\binom{n}{x}\theta^{x}(1-\theta)^{n-x},\quad x=0,1,\cdots,n,\quad 0<\theta<1$$

然后求X的边际分布

$$m(x)=\int_0^1\binom{n}{x}\theta^{x}(1-\theta)^{n-x}\,d\theta=\binom{n}{x}\frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)}$$

最后求出θ的后验分布

$$\pi(\theta|x)=\frac{h(x,\theta)}{m(x)}=\frac{\Gamma(n+2)}{\Gamma(x+1)\Gamma(n-x+1)}\theta^{(x+1)-1}(1-\theta)^{(n-x+1)-1},$$

$$0<\theta<1$$

最后的结果说明$\theta|x\sim Be(x+1,n-x+1)$，其后验期望估计为

$$\hat{\theta}_B=E(\theta|x)=\frac{x+1}{n+2}.$$

3.5.3 贝叶斯估计*

- 某些场合，贝叶斯估计要比极大似然估计更合理一点。比如：“抽检3个全是合格品”与“抽检10个全是合格品”，后者的质量比前者更信得过。这种差别在不合格品率的极大似然估计中反映不出来（两者都为0），而用贝叶斯估计两者分别是 0.2 和 0.083。
- 在极端情况下，贝叶斯估计比极大似然估计更符合人们的理念。

3.5.3 贝叶斯估计*

- **例3.25** 设$X_1, \ldots, X_n$是来自正态分布$N(\mu, {\sigma_0}^2)$的一个样本，其中${\sigma_0}^2$已知， μ未知，假设μ的先验分布亦为正态分布$N(\theta, \tau^2)$，其中先验均值θ和先验方差τ^2均已知，试求μ的贝叶斯估计。
- **解：** 样本$\boldsymbol{x}$的分布和μ的先验分布分别为

$$f(\boldsymbol{x}|\mu) = (2\pi\sigma_0^2)^{-n/2}\exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

$$\pi(\mu) = (2\pi\tau^2)^{-1/2}\exp\left\{-\frac{1}{2\tau^2}(\mu-\theta)^2\right\}$$

3.5.3 贝叶斯估计*

由此可以写出$\boldsymbol{x}$与μ的联合分布

$$h(\boldsymbol{x},\mu)=k_1\cdot\exp\left\{-\frac{1}{2}\left[\frac{n\mu^2-2n\mu\bar{x}+\sum_{i=1}^n x_i^2}{\sigma_0^2}+\frac{\mu^2-2\theta\mu+\theta^2}{\tau^2}\right]\right\}$$

其中$\bar{x}=\frac{1}{n}\sum_{i=1}^n x_i$，$k_1=(2\pi)^{-(n+1)/2}\tau^{-1}\sigma_0^{-n}$。若记

$$A=\frac{n}{\sigma_0^2}+\frac{1}{\tau^2},\quad B=\frac{n\bar{x}}{\sigma_0^2}+\frac{\theta}{\tau^2},\quad C=\frac{\sum_{i=1}^n x_i^2}{\sigma_0^2}+\frac{\theta^2}{\tau^2}$$

则有$h(\boldsymbol{x},\mu)=k_1\exp\left\{-\frac{1}{2}[A\mu^2-2B\mu+c]\right\}$

$$=k_1\exp\left\{-\frac{(\mu-B/A)^2}{2/A}-\frac{1}{2}(C-B^2/A)\right\}$$

3.5.3 贝叶斯估计*

注意到A,B,C均与μ无关，由此容易算得样本的边际密度函数

$$m(\boldsymbol{x})=\int_{-\infty}^{\infty}h(\boldsymbol{x},\mu)d\mu=k_1\exp\left\{-\frac{1}{2}(C-B^2/A)(2\pi/A)^{1/2}\right\}$$

应用贝叶斯公式即可得到后验分布

$$\pi(\mu|\boldsymbol{x})=\frac{h(\boldsymbol{x},\mu)}{m(\boldsymbol{x})}=(2\pi/A)^{1/2}\exp\left\{-\frac{1}{2/A}(\mu-B/A)^2\right\}$$

这说明在样本给定后，μ的后验分布为$N(B/A,1/A)$，即

$$\mu|\boldsymbol{x}\sim N\left(\frac{n\bar{x}\sigma_0^{-2}+\theta\tau^{-2}}{n\sigma_0^{-2}+\tau^{-2}},\frac{1}{n\sigma_0^{-2}+\tau^{-2}}\right)$$

3.5.3 贝叶斯估计*

后验均值即为其贝叶斯估计：

$$\hat{\mu} = \frac{n/\sigma_0^2}{n/\sigma_0^2 + 1/\tau^2}\bar{x} + \frac{1/\tau^2}{n/\sigma_0^2 + 1/\tau^2}\theta$$

它是样本均值$\bar{x}$与先验均值θ的加权平均。

3.5.3 贝叶斯估计*

- **补16** 某人每天早上在汽车站等公共汽车的时间（单位：min）服从均匀分布$U(0,\theta)$，其中θ未知，假设θ的先验分布为$\pi(\theta)=\begin{cases}192/\theta^4, & \theta\ge 4\\ 0, & \theta<4\end{cases}$，假如此人在三个早上等车的事件分别为5，3，8min，求θ的后验分布。

- **解：**$X_1,\dots,X_n$与θ的联合分布为

$$h(x_1,\dots,x_n,\theta)=\theta^{-n}\frac{192}{\theta^4},0<x_{(1)}<x_{(n)}<\theta,\theta\ge 4,$$

3.5.3 贝叶斯估计*

此处，$x_{(1)}=3$，$x_{(3)}=8$，所以$X_1,\dots,X_3$与θ的联合分布为

$$h(x_1,\dots,x_3,\theta)=192\theta^{-3-4},\theta>8,$$

于是θ的后验分布为

$$\pi(\theta|x_1,\dots,x_3)=\frac{192\theta^{-3-4}}{\int_8^{\infty}192\theta^{-3-4}\,d\theta}=6\times 8^6\theta^{-7}$$

$$=1572864\theta^{-7},\theta>8.$$

3.5.4 共轭先验分布*

若后验分布$\pi(\theta|\boldsymbol{x})$与$\pi(\theta)$属于同一个分布族，则称该分布族是$\theta$的**共轭先验分布(族)**。

- 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玛分布$Ga(\alpha,\lambda)$；
- 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玛分布$IGa(\alpha,\lambda)$。

§3.6 区间估计

3.6.1 区间估计的概念

- 定义3.9 设θ是总体的一个参数，其参数空间Θ，$X_1,\dots,X_n$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$\hat{\theta}_L=\hat{\theta}_L(X_1,\dots,X_n)$和$\hat{\theta}_U=\hat{\theta}_U(X_1,\dots,X_n)$，若对任意的$\theta\in\Theta$，有

$$P_\theta\left(\hat{\theta}_L\le\theta\le\hat{\theta}_U\right)\ge 1-\alpha,$$

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

3.6.1 区间估计的概念

这里置信水平$1-\alpha$有一个频率解释：

在大量重复使用θ的置信区间$[\hat{\theta}_L, \hat{\theta}_U]$时，每次得到的样本观测值是不同的，从而每次得到的区间也是不一样的。对一次具体的观测值而言，θ可能在$[\hat{\theta}_L, \hat{\theta}_U]$内，也可能不在。平均而言，在这大量的估计观测值中，至少有$100(1-\alpha)\%$包含θ。

3.6.1 区间估计的概念

- 例3.26 设$X_1, X_2, \dots, X_{10}$是来自$N(\mu, \sigma^2)$的样本，则μ的置信水平为$1-\alpha$的置信区间为

$$\left[\bar{X} - t_{1-\alpha/2}(9)S/\sqrt{10}\ , \bar{X} + t_{1-\alpha/2}(9)S/\sqrt{10}\right]$$

其中， $\bar{X}, S$ 分别为样本均值和样本标准差。这个置信区间的由来将在3.6.3节中说明，这里用它来说明置信区间与置信水平的含义。

若取$\alpha = 0.10$，则$t_{0.95}(9) = 1.8331$，上式化为

$$[\bar{X} - 0.5797S\ , \bar{X} + 0.5797S]$$

3.6.1 区间估计的概念

现假定$\mu = 15$，$\sigma^2 = 4$，则我们可以用随机模拟方法由$N(15,4)$ 产生一个容量为10的样本，如下即是这样一个样本：

14.85	13.01	13.50	14.93	16.97
13.80	17.95	13.37	16.29	12.38

由该样本可以算得$\bar{x} = 14.705$，$s = 1.843$

从而得到μ的一个区间估计为

$$[14.705 - 0.5797 \times 1.843，14.705 + 0.5797 \times 1.843] = [13.637，15.773]$$

该区间包含μ的真值—15。现重复这样的方法100次，可以得到100个样本，也就得到100个区间，我们将这100个区间画在图3.6.1上。

3.6.1 区间估计的概念

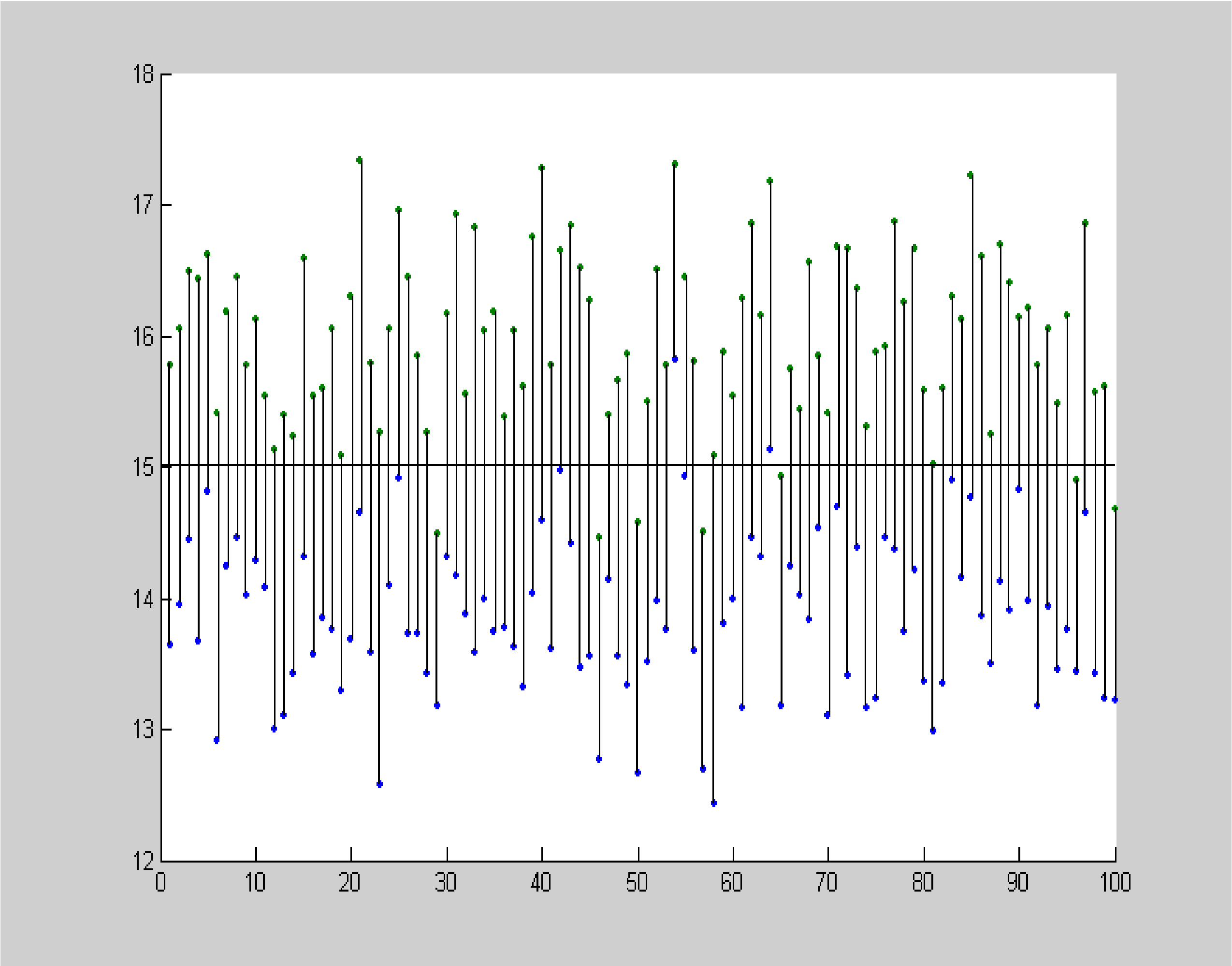

图3.6.1 μ的置信水平为0.90的置信区间

这100个区间中 有91个包含参数真值15，另外9个不包含参数真值。

3.6.1 区间估计的概念

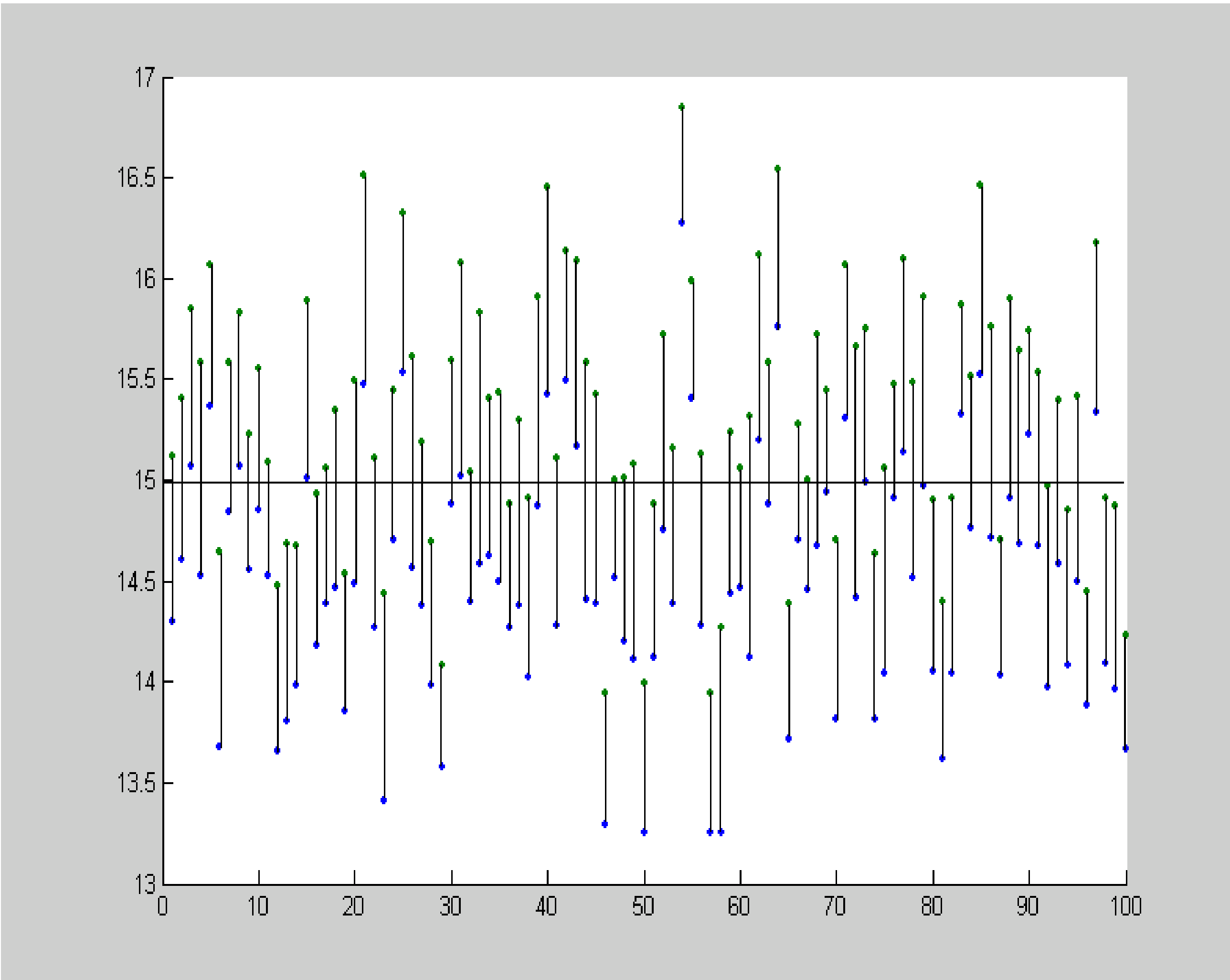

图3.6.2 μ的置信水平为0.50的置信区间

取α=0.50，我们也可以给出100个这样的区间，见左图。可以看出，这100个区间中有50个包含参数真值15，另外50个不包含参数真值。

3.6.1 区间估计的概念

- 定义3.10 沿用定义3.9的记号，如对给定的 $\alpha(0<\alpha<1)$，对任意的$\theta\in\Theta$，有

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

称$\left[\hat{\theta}_L,\hat{\theta}_U\right]$为$\theta$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。

3.6.1 区间估计的概念

- **定义3.11** 若对给定的$\alpha(0<\alpha<1)$和任意的$\theta\in\Theta$，有$P_\theta\left(\hat{\theta}_L\le\theta\right)\ge 1-\alpha$，则称$\hat{\theta}_L$为$\theta$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$\hat{\theta}_L$为$\theta$的$1-\alpha$**同等置信下限**。
- **定义3.12** 若对给定的$\alpha(0<\alpha<1)$和任意的$\theta\in\Theta$，有$P_\theta\left(\hat{\theta}_U\ge\theta\right)\ge 1-\alpha$，则$\hat{\theta}_U$为$\theta$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$\hat{\theta}_U$为$\theta$的$1-\alpha$**同等置信上限**。

单侧置信限是置信区间的特殊情形。因此，寻求置信区间的方法可以用来寻找单侧置信限。

3.6.2 枢轴量法

构造未知参数θ的置信区间的最常用的方法是**枢轴量法**，其步骤可以概括为如下三步:

1. 设法构造一个样本和θ的函数$G = G(X_1, \dots, X_n, \theta)$使得$G$的分布不依赖于未知参数。一般称具有这种性质的$G$为枢轴量。
2. 适当选择两个常数c, d，使对给定的$\alpha(0 < \alpha < 1)$有
$$P(c \le G \le d) = 1 - \alpha.$$
3. 假如能将$c \le G \le d$进行不等式等价变形化为$\hat{\theta}_L \le \theta \le \hat{\theta}_U$，则有
$$P\left(\hat{\theta}_L \le \theta \le \hat{\theta}_U\right) = 1 - \alpha,$$
这表明$[\hat{\theta}_L, \hat{\theta}_U]$是$\theta$的$1 - \alpha$同等置信区间。

3.6.2 枢轴量法

关于置信区间的构造有两点说明:

- 满足置信度要求的c与d通常不唯一。若有可能，应选平均长度$E\left(\hat{\theta}_U-\hat{\theta}_L\right)$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$E\left(\hat{\theta}_U-\hat{\theta}_L\right)$尽可能短的$c$与$d$，这往往很难实现，因此，常这样选择$c$与$d$，使得两个尾部概率各为$\alpha/2$，即$P(G<c)=P(G>d)=\alpha/2$，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。

3.6.2 枢轴量法

- 例3.27 设$X_1,\dots,X_n$是来自均匀总体$U(0,\theta)$的一个样本，试对给定的$\alpha(0<\alpha<1)$给出θ的$1-\alpha$同等置信区间。
- **解:**

（1）取$X_{(n)}/\theta$作为枢轴量，其密度函数为

$$f(y;\theta)=ny^{n-1}，\ 0<y<1;$$

（2）$X_{(n)}/\theta$的分布函数为$F(y)=y^n,0<y<1$，故$P(c\le x_{(n)}/\theta\le d)=d^n-c^n$，因此我们可以适当地选择$c$和$d$满足

$$d^n-c^n=1-\alpha。$$

3.6.2 枢轴量法

（3）利用不等式变形可容易地给出θ的$1-\alpha$同等置信区间为$[X_{(n)}/d, X_{(n)}/c]$，该区间的平均长度为$\left(\frac{1}{c}-\frac{1}{d}\right)EX_{(n)}$。不难看出，在$0\le c<d\le 1$及$d^n-c^n=1-\alpha$的条件下，当$d=1$，$c=\sqrt[n]{\alpha}$时，$\frac{1}{c}-\frac{1}{d}$取得最小值，这说明$[X_{(n)}, X_{(n)}/\sqrt[n]{\alpha}]$是$\theta$的置信水平为$1-\alpha$最短置信区间。

3.6.2 枢轴量法

- **补17***设总体X的密度函数为$\lambda e^{-\lambda x}I_{[x>0]}$，其中$\lambda > 0$为未知参数，$X_1, X_2, \dots, X_n$抽自总体的简单随机样本，求$\lambda$的置信水平为$1-\alpha$的置信区间。

- **解：** 由指数分布和伽马分布的关系知

$\sum_{i=1}^{n} X_i \sim Ga(n, \lambda)$，根据伽马分布的性质，

$$2\lambda \sum_{i=1}^{n} X_i \sim Ga\left(n, \frac{1}{2}\right) = \chi^2(2n),$$

3.6.2 枢轴量法

$$P\left(\chi^2_{\frac{\alpha}{2}}(2n) \le 2\lambda \sum_{i=1}^{n} X_i \le \chi^2_{1-\frac{\alpha}{2}}(2n)\right) = 1-\alpha$$

因此可得λ的置信水平为$1-\alpha$的置信区间为

$$\left[\frac{\chi^2_{\alpha/2}(2n)}{2n\bar{X}}, \frac{\chi^2_{1-\alpha/2}(2n)}{2n\bar{X}}\right].$$

3.6.3 单个正态总体参数的置信区间

一、σ已知时μ的置信区间

- 在这种情况下，枢轴量可选为$G=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$，$c$和$d$应满足$P(c\le G\le d)=\Phi(d)-\Phi(c)=1-\alpha$，经过不等式变形可得

$$P_\mu(\bar{X}-d\sigma/\sqrt{n}\le\mu\le\bar{X}-c\sigma/\sqrt{n})=1-\alpha$$

该区间长度为$(d-c)\sigma/\sqrt{n}$。当$d=-c=u_{1-\alpha/2}$时，$d-c$达到最小，由此给出了的同等置信区间为

$$\left[\bar{X}-u_{1-\alpha/2}\sigma/\sqrt{n},\ \bar{X}+u_{1-\alpha/2}\sigma/\sqrt{n}\right]$$

这是一个$\bar{X}$以为中心，半径为$u_{1-\alpha/2}\sigma/\sqrt{n}$的对称区间，常将之表示为$\bar{X}\pm u_{1-\alpha/2}\sigma/\sqrt{n}$。

3.6.3 单个正态总体参数的置信区间

- **例3.28** 用天平秤某物体的重量9次，得平均值为 $\bar{x}=15.4$（克），已知天平秤量结果为正态分布，其标准差为0.1克。试求该物体重量的0.95置信区间。
- **解：** 此处$1-\alpha=0.95$，$\alpha=0.05$，查表知$u_{0.975}=1.96$，于是该物体重量μ的0.95置信区间为

$$\bar{x}\pm u_{1-\alpha/2}\sigma/\sqrt{n}=15.4\pm 1.96\times 0.1/\sqrt{9}$$

$$=15.4\pm 0.0653,$$

从而该物体重量的0.95置信区间为 [15.3347，15.4653] 。

3.6.3 单个正态总体参数的置信区间

- 例3.29 设总体为正态分布$N(\mu,1)$，为得到μ的置信水平为0.95的置信区间长度不超过1.2，样本容量应为多大?

- **解：** 由题设条件知μ的0.95置信区间为

$$\left[\bar{X}-u_{1-\alpha/2}/\sqrt{n}，\ \bar{X}+u_{1-\alpha/2}/\sqrt{n}\right]$$

其区间长度为$2u_{1-\alpha/2}/\sqrt{n}$，它仅依赖于样本容量n而与样本具体取值无关。

现要求$2\,u_{1-\alpha/2}/\sqrt{n}\le 1.2$，立即有$n\ge(2/1.2)^2u^2{}_{1-\alpha/2}$。

现$1-\alpha=0.95$，故$u_{1-\alpha/2}=1.96$，从而

$$n\ge(5/3)^2 1.96^2=10.67\approx 11。$$

即样本容量至少为11时才能使得μ的置信水平为0.95的置信区间长度不超过1.2。

3.6.3 单个正态总体参数的置信区间

- 补18 0.50，1.25，0.80，2.00是取自总体X的样本，已知$Y=\ln X$服从正态分布$N(\mu,1)$，求μ的置信水平为0.95的置信区间。

- **解：** 将数据进行对数变换，得到$Y=\ln X$的样本值为

-0.6931，0.2231，-0.2231.0.6931.

它可看作是来自正态总体$N(\mu,1)$的样本，其样本均值为$\bar{y}=0$，由于$\sigma=1$已知，因此，μ的置信水平为0.95的置信区间为

$$[\bar{y}-u_{1-\frac{\alpha}{2}}/\sqrt{n},\bar{y}+u_{1-\frac{\alpha}{2}}/\sqrt{n}]=[-0.9800,0.9800].$$

3.6.3 单个正态总体参数的置信区间

二、 σ未知时μ的置信区间

- 这时可用T统计量，因为$T=\frac{\sqrt{n}(\bar{X}-\mu)}{S}\sim t(n-1)$，因此$T$可以用来作为枢轴量。完全类似于上一小节，可得到$\mu$的$1-\alpha$置信区间为

$$\left[\bar{X}-t_{1-\alpha/2}(n-1)S/\sqrt{n},\ \bar{X}+t_{1-\alpha/2}(n-1)S/\sqrt{n}\right]$$

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

3.6.3 单个正态总体参数的置信区间

- **例3.30** 假设轮胎的寿命服从正态分布。为估计某种轮胎的平均寿命，现随机地抽12只轮胎试用，测得它们的寿命（单位：万千米）如下：

  4.68 4.85 4.32 4.85 4.61 5.02

  5.20 4.60 4.58 4.72 4.38 4.70

  试求平均寿命的0.95置信区间。

- **解：** 此处正态总体标准差未知，可使用t分布求均值的置信区间。经计算$\bar{x} = 4.7092$，$s^2 = 0.0615$。取$\alpha = 0.05$，查表知$t_{0.975}(11) = 2.2010$，于是平均寿命的0.95置信区间为（单位：万千米）

$$4.7092 \pm 2.2010 \cdot \sqrt{0.0615}/\sqrt{12} = [4.5516，4.8668]$$

3.6.3 单个正态总体参数的置信区间

在实际问题中，由于轮胎的寿命越长越好，因此可以只求平均寿命的置信下限，也即构造单边的置信下限。由于

$$P\left(\frac{\sqrt{n}(\bar{X}-\mu)}{S} < t_{1-\alpha}(n-1)\right) = 1-\alpha$$

由不等式变形可知μ的$1-\alpha$置信下限为

$$\bar{X} - t_{1-\alpha}(n-1)S/\sqrt{n}$$

将$t_{0.95}(11) = 1.7959$代入计算可得平均寿命μ的0.95置信下限为4.5806（万千米）。

3.6.3 单个正态总体参数的置信区间

- **补19** 为了估计1分钟1次广告的平均费用，抽出了15个电视台的随机样本。样本的平均值$\bar{x}=2000$，其标准差$s=1000$元。假定所有被抽样的这类电视台近似服从正态分布，试构造总体平均值置信水平为95%的置信区间。

- **解：** 总体近似服从正态分布，但方差未知，$\bar{x}=2000$，$s=1000$，$n=15$。为了求概率度，查自由度$n-1=14$的t分布表，与置信水平95%相应的t值为：

$$t_{1-\alpha/2}(n-1)=t_{0.025}(14)=2.145$$

3.6.3 单个正态总体参数的置信区间

置信区间为:

$$\bar{x} \pm t_{1-\alpha/2}(n-1)s/\sqrt{n}$$

$$2000 \pm 2.145 \times 1000/\sqrt{15}$$

即

（1446.16，2553.84）

显然我们有95%的把握说明，总体平均数处在1446.16~2553.84元之间。

3.6.3 单个正态总体参数的置信区间

三、σ^2的置信区间

- 取枢轴量$G=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$，由于$\chi^2$分布是偏态分布，寻找平均长度最短区间很难实现，一般都用等尾置信区间：采用χ^2的两个分位数$\chi^2_{\alpha/2}(n-1)$和$\chi^2_{1-\alpha/2}(n-1)$，在χ^2分布两侧各截面积为$\alpha/2$的部分，使得

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

由此给出σ^2的$1-\alpha$置信区间为

$$\left[(n-1)\,S^2/\chi^2_{1-\alpha/2}(n-1)，(n-1)\,S^2/\chi^2_{\alpha/2}(n-1)\right]$$

3.6.3 单个正态总体参数的置信区间

- **例3.31** 某厂生产的零件重量服从正态分布$N(\mu,\sigma^2)$，现从该厂生产的零件中抽取9个，测得其重量为（单位：克）

  45.3 45.4 45.1 45.3 45.5 45.7 45.4 45.3 45.6

  试求总体标准差σ的0.95置信区间。

- **解：** 由数据可算得$s^2=0.0325$，$(n-1)s^2=8\times 0.0325=0.26$。查表知$\chi^2_{0.025}(8)=2.1797$，$\chi^2_{0.975}(8)=17.5345$，代入可得$\sigma^2$的0.95置信区间为

$$\left[\frac{0.26}{17.5345},\ \frac{0.26}{2.1797}\right]=[0.0148,\ 0.1193]$$

从而σ的0.95置信区间为: $[0.1218,\ 0.3454]$。

3.6.3 单个正态总体参数的置信区间

- 补20 某制造厂的一名生产管理人员需要知道完成某件工作所需的时间。为此他进行了一项研究，得出一个适于分析的31个观察值组成的随机样本，从样本数据算出的方差为0.3小时。计算：

（1）构造σ^2的95%的置信区间；

（2）构造σ的95%的置信区间；

（3）构造置信区间时作了何种假定？

- **解：**　（1）σ^2的$1-\alpha$置信区间为

$$\left[(n-1)S^2/\chi^2_{1-\alpha/2}(n-1)\text{，}(n-1)S^2/\chi^2_{\alpha/2}(n-1)\right]$$

由题意可知，$s^2=0.3$，自由度$=\mathrm{n}-1=31-1=30$，查自由度为30的χ^2分布表得：

3.6.3 单个正态总体参数的置信区间

$$\chi^2_{\frac{\alpha}{2}}(30) = \chi^2_{\frac{0.05}{2}}(30) = \chi^2_{0.025}(30) = 46.979,$$

$$\chi^2_{1-\frac{\alpha}{2}}(30) = \chi^2_{1-\frac{0.05}{2}}(30) = \chi^2_{0.975}(30) = 16.791。$$

代入公式得：

$$\frac{(31-1)\times 0.3}{46.979} < \sigma^2 < \frac{(31-1)\times 0.3}{16.791}$$

$$0.1916 < \sigma^2 < 0.5360$$

则有95%的把握说σ^2落在0.1916~0.5360之间的范围内。

（2）其总体标准差的置信区间为：$0.4377 < \sigma < 0.7321$。

（3）被抽样的总体服从或近似服从正态分布是置信区间估计的假定条件。

3.6.3 单个正态总体参数的置信区间

- **补21** 用一个仪表测量某一物理量9次，得样本均值$\bar{x}=56.32$，样本标准差$s=0.22$.测量标准差σ大小反映了测量仪表的精度，试求σ置信水平为0.95的置信区间。

- **解：** σ^2的$1-\alpha$置信区间为

$$\left[(n-1)S^2/\chi^2_{1-\alpha/2}(n-1)，(n-1)S^2/\chi^2_{\alpha/2}(n-1)\right]$$

由题意得$(n-1)s^2=8\times 0.22^2=0.3872$，查表知$\chi^2_{0.025}(8)=2.1797, \chi^2_{0.975}(8)=17.5345$，故$\sigma^2$的95%置信区间为

$$\left[\frac{0.3872}{17.5345},\frac{0.3872}{2.1797}\right]=[0.0221,0.1776],$$

从而σ的置信水平为0.95的置信区间为$[0.1487,0.4215]$.

3.6.4 大样本置信区间

- 在有些场合，寻找枢轴量及其分布比较困难。在样本容量充分大时，可以用渐近分布来构造近似的置信区间。一个典型的例子是关于比例p的置信区间。

3.6.4 大样本置信区间

设$X_1,\cdots,X_n$是来自$b(1,p)$的样本,有

$$U=\frac{\bar{X}-p}{\sqrt{p(1-p)/n}}\dot{\sim}N(0,1)$$

对给定α,

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

通过变形，可得到置信区间为

$$\left[\frac{1}{1+\frac{\lambda}{n}}\left(\bar{X}+\frac{\lambda}{2n}-\sqrt{\frac{\bar{X}(1-\bar{X})\lambda}{n}+\frac{\lambda^2}{4n^2}}\right),\ \frac{1}{1+\frac{\lambda}{n}}\left(\bar{X}+\frac{\lambda}{2n}+\sqrt{\frac{\bar{X}(1-\bar{X})\lambda}{n}+\frac{\lambda^2}{4n^2}}\right)\right]$$

其中记$\lambda=u_{1-\alpha/2}^2$，实用中通常略去$\frac{\lambda}{n}$项，于是可将置信区间近似为

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

3.6.4 大样本置信区间

- 例3.32 对某事件A作120次观察，A发生36次。试给出事件A发生概率p的0.95置信区间。
- **解：** 此处$n = 120$，$\bar{x} = 36/120 = 0.3$而$u_{0.975} = 1.96$，于是p的0.95（双侧）置信下限和上限分别为

$$\hat{p}_L = 0.3 - 1.96 \times \sqrt{\frac{0.3 \times 0.7}{120}} = 0.218$$

$$\hat{p}_U = 0.3 + 1.96 \times \sqrt{\frac{0.3 \times 0.7}{120}} = 0.382$$

故所求的置信区间为[0.218，0.382]。

3.6.4 大样本置信区间

- 补22 在一批货物中随机抽取80件，发现有11件不合格品，试求这批货物的不合格品率的置信水平为0.90的置信区间。

- **解：** 此处$n=80$较大，可用正态分布求其近似置信区间。不合格品率的$1-\alpha$近似置信区间为

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

3.6.4 大样本置信区间

此处$\bar{x}=\frac{11}{80}=0.1375, u_{0.95}=1.645$，因而不合格品率的置信水平为0.90的置信区间为

$$\left[0.1375-1.645\sqrt{0.1375\times\frac{0.8625}{80}},0.1375+1.645\sqrt{0.1375\times\frac{0.8625}{80}}\right]$$

$$=[0.0742,0.2008].$$

3.6.5 样本量的确定

- **例3.33** 某传媒公司欲调查电视台某综艺节目收视率p，为使得p的$1-\alpha$置信区间长度不超过$2d_0$，问应调查多少用户？
- **解：** 这是关于二点分布比例p的置信区间问题，由（3.6.11）知，$1-\alpha$的置信区间长度为$2u_{1-\alpha/2}\sqrt{\bar{X}(1-\bar{X})/n}$，这是一个随机变量，但$\bar{X}\in(0,1)$，所以对任意的观测值有$\bar{X}(1-\bar{X})\le 0.5^2=0.25$。这也就是说$p$的$1-\alpha$的置信区间长度不会超过$u_{1-\alpha/2}/\sqrt{n}$。现要求$p$的置信区间长度不超过$2d_0$，只需要$u_{1-\alpha/2}/\sqrt{n}\le 2d_0$即可，从而

$$n\ge\left(\frac{u_{1-\alpha/2}}{2d_0}\right)^2.$$

3.6.5 样本量的确定

这是一类常见的寻求样本量的问题。比如，若取 $d_0 = 0.02$，$\alpha = 0.05$，则

$$n \geq \left(\frac{u_{0.975}}{0.04}\right)^2 = \left(\frac{1.96}{0.04}\right)^2 = 2401$$

这表明，要使综艺节目收视率p的0.95置信区间的半径不超过0.02，则需要对2401个用户作调查。

3.6.6 两个正态总体下的置信区间

- 设$X_1,\cdots,X_m$是来自$N(\mu_1,{\sigma_1}^2)$的样本， $Y_1,\cdots,Y_n$是来自$N(\mu_2,{\sigma_2}^2)$的样本，且两个样本相互独立。$\bar{X}$与$\bar{Y}$ 分别是它们的样本均值，

$S_X^2=\frac{1}{m-1}\sum_{i=1}^{m}(X_i-\bar{X})^2$和$S_Y^2=\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2$分别是它们的样本方差。下面讨论两个均值差和两个方差比的置信区间。

3.6.6 两个正态总体下的置信区间

一、$\mu_1-\mu_2$的置信区间

1、${\sigma_1}^2$和${\sigma_2}^2$已知时的两样本u区间

$$\left[\bar{X}-\bar{Y}-u_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}},\ \bar{X}-\bar{Y}+u_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}\right]$$

2、${\sigma_1}^2={\sigma_2}^2={\sigma}^2$未知时的两样本$t$区间

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

其中$S_w^2=\frac{(m-1)S_X^2+(n-1)S_Y^2}{m+n-2}$.

3.6.6 两个正态总体下的置信区间

3、$\sigma_2{}^2/\sigma_1{}^2 = c$已知时的两样本$t$区间

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

其中$S_w^2=\frac{(m-1)S_X^2+(n-1)S_Y^2/c}{m+n-2}$.

3.6.6 两个正态总体下的置信区间

- 补23某银行负责人想知道存户存入2家银行的钱数，他从每一家银行各抽选了1个由25个存户组成的随机样本。样本平均值如下：银行A：$\bar{x}_A = 450$元；银行B：$\bar{x}_B = 325$元。2个总体均服从方差分别为$\sigma_A^2 = 750$和$\sigma_B^2 = 850$的正态分布。试构造$\mu_A - \mu_B$的95%的置信区间。
- **解：** 由于2个总体均服从正态分布，因此$\bar{X}_A - \bar{X}_B$也服从正态分布，从而计算总体均值之差的置信区间可用：

$$\left[\bar{X}_A - \bar{X}_B - u_{1-\alpha/2}\sqrt{\frac{\sigma_A^2}{m} + \frac{\sigma_B^2}{n}},\ \bar{X}_A - \bar{X}_B + u_{1-\alpha/2}\sqrt{\frac{\sigma_A^2}{m} + \frac{\sigma_B^2}{n}}\right]$$

已知$\sigma_A^2 = 750$，$\sigma_B^2 = 850$，$\bar{x}_A = 450$，$\bar{x}_B = 325$，

3.6.6 两个正态总体下的置信区间

$u_{1-\alpha/2}$=1.96，所以95%的置信区间为：

$$\left[(450-325)-1.96\times\sqrt{\frac{750}{25}+\frac{850}{25}},(450-325)+1.96\times\sqrt{\frac{750}{25}+\frac{850}{25}}\right]$$

即（109.32，140.68），这就意味着有95%的把握认为总体均值之差在109.32~140.68元之间。

3.6.6 两个正态总体下的置信区间

- **例3.34** 为比较两个小麦品种的产量，选择18块条件相似的试验田，采用相同的耕作方法作试验，结果播种甲品种的8块试验田的亩产量和播种乙品种的10块试验田的亩产量（单位：千克/亩）分别为：

甲品种 628 583 510 554 612 523 530 615

乙品种 535 433 398 470 567 480 498 560 503 426

假定亩产量均服从正态分布，试求这两个品种平均亩产量差的置信区间($\alpha = 0.05$)。

3.6.6 两个正态总体下的置信区间

- **解：** 设甲品种平均亩产量为μ_1，乙品种平均亩产量为μ_2。以$X_1, \ldots, X_8$记甲品种的亩产量，$Y_1, \ldots, Y_{10}$记乙品种的亩产量，由样本数据可计算得到

$$\bar{x} = 569.38，\ s_x^2 = 2140.55，\ m = 8$$

$$\bar{y} = 487.00，\ s_y^2 = 3256.22，\ n = 10$$

下面分两种情况讨论。

3.6.6 两个正态总体下的置信区间

(1) 若已知两个品种亩产量的标准差相同，则可采用两样本t区间。$\mu_1-\mu_2$的 $1-\alpha$ 置信区间为

$$\bar{X}-\bar{Y}\pm\sqrt{\frac{m+n}{mn}}S_w t_{1-\alpha/2}(m+n-2),$$

其中$S_w=\sqrt{\frac{(m-1)S_X^2+(n-1)S_Y^2}{m+n-2}}$.

3.6.6 两个正态总体下的置信区间

代入数据得：

$$s_w=\sqrt{\frac{(m-1)s_x^2+(n-1)s_y^2}{m+n-2}}=\sqrt{\frac{7\times 2140.55+9\times 3256.22}{16}}=52.6129$$

$$t_{1-\alpha/2}(m+n-2)=t_{0.975}(16)=2.1199$$

$$t_{1-\alpha/2}(m+n-2)s_w\sqrt{\frac{1}{m}+\frac{1}{n}}=2.1199\times 52.6129\times\sqrt{\frac{1}{8}+\frac{1}{10}}=52.91$$

故$\mu_1-\mu_2$的0.95置信区间为

$$[569.38-487-52.91，569.38-487+52.91]=[29.47，135.29]$$

3.6.6 两个正态总体下的置信区间

(2) 若两个品种亩产量的方差不等，则可采用近似t区间。

$\mu_1-\mu_2$的 $1-\alpha$ 置信区间为

$$\bar{X}-\bar{Y}\pm S_0 t_{1-\alpha/2}(l)$$

其中$S_0^2=S_X^2/m+S_Y^2/n$,$l=\dfrac{S_0^4}{\frac{S_X^4}{m^2(m-1)}+\frac{S_Y^4}{n^2(n-1)}}$.

3.6.6 两个正态总体下的置信区间

代入数据得

$$s_0^2 = 2140.55/8+3256.22/10=593.19,$$

$$s_0 = 24.36,$$

$$l = \frac{593.19^{2}}{\dfrac{2140.55^2}{8^2 \times 7} + \dfrac{3256.22^2}{10^2 \times 11}} = 15.99 \approx 16,$$

$$s_0 t_{0.975}(l) = 24.36 \times 2.1199 = 51.64,$$

于是$\mu_1 - \mu_2$的0.95近似置信区间为

$$[30.74，134.02]$$

3.6.6 两个正态总体下的置信区间

二、$\sigma_1{}^2/\sigma_2{}^2$的置信区间

由于$(m-1)S_X^2/\sigma_1^2 \sim \chi^2(m-1)$，$(n-1)S_Y^2/\sigma_2^2 \sim \chi^2(n-1)$，且$S_X^2$与$S_Y^2$相互独立，故可仿照$F$变量构造如下枢轴量

$F=\dfrac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2} \sim F(m-1,n-1)$,对给定的$1-\alpha$，由

$$P\left(F_{\alpha/2}(m-1,n-1) \le \frac{S_X^2}{S_Y^2}\cdot\frac{\sigma_2^2}{\sigma_1^2} \le F_{1-\alpha/2}(m-1,\ n-1)\right)=1-\alpha$$

经不等式变形即给出σ_1^2/σ_2^2的如下的置信区间

$$\left[\frac{S_X^2}{S_Y^2}\cdot\frac{1}{F_{1-\alpha/2}(m-1,\ n-1)},\ \frac{S_X^2}{S_Y^2}\cdot\frac{1}{F_{\alpha/2}(m-1,\ n-1)}\right]$$

3.6.6 两个正态总体下的置信区间

- **例3.35** 某车间有两台自动机床加工一类套筒，假设套筒直径服从正态分布。现在从两个班次的产品中分别检查了5个和6个套筒，得其直径数据如下（单位：厘米）：

甲班：5.06 5.08 5.03 5.00 5.07

乙班：4.98 5.03 4.97 4.99 5.02 4.95

试求两班加工套筒直径的方差比 $\sigma_{甲}^2/\sigma_{乙}^2$ 的0.95置信区间。

解： 对给定的$1-\alpha$，σ_1^2/σ_2^2的置信区间为

$$\left[\frac{S_X^2}{S_Y^2}\cdot\frac{1}{F_{1-\alpha/2}(m-1,\ n-1)},\ \frac{S_X^2}{S_Y^2}\cdot\frac{1}{F_{\alpha/2}(m-1,\ n-1)}\right]$$

3.6.6 两个正态总体下的置信区间

代入数据得

$$F_{0.025}(4,5)=\frac{1}{F_{0.975}(5,4)}=\frac{1}{9.36}=0.1068,$$

$$F_{0.975}(4,5)=7.39,$$

由数据算得$s_{甲}^2=0.00037, s_{乙}^2=0.00092$，故置信区间为

$$[0.0544，3.7657].$$

3.6.6 两个正态总体下的置信区间

- 补24 假设人体身高服从正态分布，今抽测甲、乙两地区18岁~25岁女青年身高得数据如下：甲地区抽取10名，样本均值1.64m，样本标准差0.2m；乙地区抽取10名，样本均值1.62m，样本标准差0.4m。求：

（1）两正态总体方差比的置信水平为95%的置信区间；

（2）两正态总体均值差的置信水平为95%的置信区间。

- **解：** 设$X_1,\dots,X_{10}$为甲地区抽取的女青年身高，$Y_1,\dots,Y_{10}$为乙地区抽取的女青年身高，由题设条件：

$$\bar{x}=1.64, s_x=0.2, \bar{y}=1.62, s_y=0.4.$$

3.6.6 两个正态总体下的置信区间

（1）σ_1^2/σ_2^2的$1-\alpha$的置信区间为

$$\left[\frac{S_X^2}{S_Y^2}\cdot\frac{1}{F_{1-\alpha/2}(m-1,\ n-1)},\ \frac{S_X^2}{S_Y^2}\cdot\frac{1}{F_{\alpha/2}(m-1,\ n-1)}\right]$$

此处$\alpha=0.05, m=n=10$，查表得

$$F_{0.975}(9,9)=4.03\ ,\ F_{0.025}=\frac{1}{F_{0.975}(9,9)}=\frac{1}{4.03},$$

由此，σ_1^2/σ_2^2的置信水平为95%的置信区间为

$$\left[\frac{0.2^2}{0.4^2}\cdot\frac{1}{4.03},\frac{0.2^2}{0.4^2}\cdot 4.03\right]=[0.0620,1.0075].$$

3.6.6 两个正态总体下的置信区间

（2）由（1），σ_1^2/σ_2^2 的置信水平为95%的置信区间包含1，因此有一定理由假定两个正态总体的方差相等，此时 $\mu_1-\mu_2$的 $1-\alpha$ 置信区间为

$$\bar{X}-\bar{Y}\pm\sqrt{\frac{m+n}{mn}}S_w t_{1-\alpha/2}(m+n-2)$$

其中$S_w=\sqrt{\frac{(m-1)S_X^2+(n-1)S_Y^2}{m+n-2}}$.

3.6.6 两个正态总体下的置信区间

代入数据得

$$s_w^2=\frac{(m-1)s_x^2+(n-1)s_y^2}{m+n-2}=\frac{9\times 0.2^2+9\times 0.4^2}{10+10-2}=0.1$$

查表得$t_{0.975}(18)=2.1009$，故两正态总体均值差的置信水平为95%的置信区间为

$$\left[1.64-1.62-2.1009\sqrt{0.1}\sqrt{\frac{10+10}{10\times 10}},1.64-1.62+2.1009\sqrt{0.1}\sqrt{\frac{10+10}{10\times 10}}\right]$$

$=[-0.2771,0.3171].$

3.6.6 两个正态总体下的置信区间

还有另一种解法就是不对方差相等作假定，而采用近似方法求均值差的置信区间，$\mu_1 - \mu_2$的 $1-\alpha$ 置信区间为

$$\bar{X} - \bar{Y} \pm S_0 t_{1-\alpha/2}(l)$$

其中$S_0^2 = S_X^2/m + S_Y^2/n$,$l = \dfrac{S_0^4}{\frac{S_X^4}{m^2(m-1)} + \frac{S_Y^4}{n^2(n-1)}}$.

3.6.6 两个正态总体下的置信区间

由于

$$s_0^2=\frac{s_x^2}{m}+\frac{s_y^2}{n}=\frac{0.04}{10}+\frac{0.16}{10}=0.02, l=\frac{0.02^2}{\frac{0.04^2}{900}+\frac{0.16^2}{900}}=13,$$

查表知$t_{0.975}(13)=2.1604$，从而两正态总体的均值差的置信水平为95%的近似置信区间为

$$\left[1.64-1.62-2.1604\sqrt{0.02}, 1.64-1.62+2.1604\sqrt{0.02}\right]$$
$$=[-0.2855,0.3255].$$

这两个置信区间相差不算太小，所以在应用中条件“方差相等”是否成立是要加以考证的。

Thank You !