第 11 周 统计的基本概念与点估计

一、数理统计的基本概念

1.1 总体和样本

- 研究对象的某个（某些）数量指标；
- 简单随机样本：随机性+独立性

十分之一原则（放回抽样、不放回抽样）

- 样本的分布函数：总体为$F(x)$的样本（容量为n）$X_1,X_2,\cdots,X_n$的观测值为$x_1,\cdots,x_n$，则此样本的分布函数为$F(x_1,\cdots,x_n)=\prod_{i=1}^{n}F(x_i)$。

1.2 统计量和经验分布函数

★ 统计量；常用统计量；顺序统计量和经验分布函数

统计量：样本的函数$T=T(X_1,X_2,\cdots,X_n)$（不含任何未知参数）

★ 常见统计量：

样本均值 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i,\quad E\bar{X}=\mu=EX$；$D(\bar{X})=\frac{\sigma^2}{2}$

样本方差$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2,\quad ES^2=\sigma^2=D(X)$

样本矩：$M_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k,\quad EM_k=EX^k\triangleq\mu_k$

样本修正方差$S^{*2}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2=M_2-M_1^2, E(S^{*2})=\frac{n-1}{n}\sigma^2$

渐近性（对应后续的相合估计）：

（1）$\overline{X}\xrightarrow{P}\mu$；（2）$S^2(or\ S^{*2})\xrightarrow{P}\sigma^2$；（3）$M_k\xrightarrow{P}\mu_k\triangleq EM_k$.

★ 顺序统计量：把样本$X_1,\cdots,X_n$按观测值从小到大的顺序将它们重新排列成

$X_{(1)}\le X_{(2)}\le\cdots\le X_{(n)}$，称统计量$X_{(1)},X_{(2)},\cdots,X_{(n)}$为顺序统计量，称

$R = X_{(n)} - X_{(1)} \equiv \max\{X_1,\cdots,X_n\} - \min\{X_1,\cdots,X_n\}$为样本极差，称

$$X_{med} = \begin{cases} X_{(\frac{n+1}{2})}, & n\text{为奇数} \\ \frac{1}{2}[X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}], & n\text{为偶数} \end{cases}$$

为样本中位数。

★ 经验分布函数：$F_n(x) = \dfrac{V_n(x)}{n} = \dfrac{1}{n}$ **{** $X_1,\cdots,X_n$ 中小于或等于 $\boldsymbol{x}$ 的个数**}**

$$= \begin{cases} 0, & \text{当} x < X_{(1)}, \\ \frac{k}{n}, & \text{当} X_{(k)} \le x < X_{(k+1)}, k = 1,2,\cdots,n-1, \\ 1, & \text{当} X_{(n)} \le x \end{cases}$$

显然$V_n(x) \sim B(n, F(x))$，从而

$$E(F_n(x)) = F(x), D(F_n(x)) = \frac{F(x)(1-F(x))}{n}$$

$$\sqrt{n}[F_n(x) - F(x)] \xrightarrow{D} N(0, F(x)(1-F(x)))$$

Thm.（Glivenko）： $\sup\limits_{x\in R} |F_n(x) - F(x)| \to 0. \quad a.s.$

★ 经验分布函数的观测值：$\tilde{F}_n(x) = \dfrac{1}{n}$ **{** $x_1,\cdots,x_n$ 中小于或等于 $\boldsymbol{x}$ 的个数**}**

注：利用这个可以画图近似给出总体的分布函数。

1.3 抽样分布

统计量的分布

★ 顺序统计量的分布：假设总体的分布密度为 $f(x)$，分布函数为$F(x)$，则

（1）$(X_{(1)}, X_{(2)}, \cdots, X_{(n)}) \sim$

$$f_{X_{(1)},X_{(2)},\cdots,X_{(n)}}(x_{(1)}, x_{(2)}, \cdots, x_{(n)}) = n!\prod_{i=1}^{n} f(x_{(i)}), \forall x_{(1)} < x_{(2)} < \cdots < x_{(n)}$$

注 1：这个可直接用随机向量的函数的分布直接给出（注意置换共有$n!$

个）。

注 2：利用这个结果，理论上可以得到其任意维边缘的分布（积分即可）

以下用启发式求边缘：

（2） $X_{(k)} \sim \dfrac{n!}{(k-1)!(n-k)!}[F(x)]^{k-1}[1-F(x)]^{n-k} f(x)$

例：均匀总体$U[0,1]$时：

$$X_{(k)} \sim \frac{n!}{(k-1)!(n-k)!} x^{k-1}[1-x]^{n-k}, 0<x<1.(i.e. \quad \sim Beta(k,n-k+1))$$

$$EX_{(k)} = \frac{k}{n+1}; DX_{(k)} = \frac{k(n-k+1)}{(n+1)^2(n+2)}$$

$$EX_{(1)} = \frac{1}{n+1}, EX_{(n)} = \frac{n}{n+1}$$

注：均匀总体$U[0,1]$，$1-X_{(n)} \overset{D}{=} X_{(1)}$ ，

$DX_{(1)} = DX_{(n)} = \dfrac{n}{(n+1)^2(n+2)}$ （**Beta 分布的方差**）

（3）

$$(X_{(i)}, X_{(j)}) \sim$$

$$\frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F(x)]^{i-1}[F(y)-F(x)]^{j-i-1}[1-F(y)]^{n-j} f(x) f(y), x \le y$$

特别有：

$$(X_{(1)}, X_{(n)}) \sim n(n-1)[F(y)-F(x)]^{n-2} f(x) f(y), x \le y$$

建议直接用启发式证明；

例：样本极差的分布密度

$R = X_{(n)} - X_{(1)}$

$$f_R(r) = \int_{-\infty}^{\infty} f_{(X_{(1)}, X_{(n)})}(u, r+u)du$$

$$= \int_{-\infty}^{\infty} n(n-1)[F(r+u) - F(u)]^{n-2} f(u) f(r+u)du, r > 0$$

特别。若 $X \sim U(0,1)$ 则对 $0 \le r \le 1$，有

$$f_R(r) = \int_{-\infty}^{\infty} n(n-1)[F(r+u) - F(u)]^{n-2} f(u) f(r+u)du$$

$$= \int_{0}^{1-r} n(n-1)[r+u-u]^{n-2} du \quad (\because 0 < r+u < 1, 0 < u < 1)$$

$$= n(n-1)r^{n-2}(1-r)$$

此时，$R \sim Beta(n-1, 2)$

$$E(R) = \frac{n-1}{n+1}, D(R) = \frac{2(n-1)}{(n+1)^2(n+2)}$$

思考：均匀总体 $U[0,1]$ 时，求 $E(X_{(n)} \mid X_{(1)})$

注：（2）和（3）严格证明举例（比较麻烦，上课不讲了），只证（1）：

$P(X_{(k)} \le x) = P$ **{** $X_1, \cdots, X_n$ 中至少有 k 个小于或等于 $\boldsymbol{x}$**}**

$$= \sum_{i=k}^{n} C_n^i [F(x)]^i [1 - F(x)]^{n-i}, k = 1, 2, \cdots, n$$

$$f_{X_{(i)}}(x)=\frac{d}{dx}P(X_{(k)}\le x)=\frac{d}{dx}\sum_{i=k}^{n}C_n^i[F(x)]^i[1-F(x)]^{n-i}$$

$$=\sum_{i=k}^{n}C_n^i\{i[F(x)]^{i-1}[1-F(x)]^{n-i}f(x)-(n-i)[F(x)]^i[1-F(x)]^{n-i-1}f(x)\}$$

$$=\sum_{i=k}^{n}C_n^i[F(x)]^{i-1}[1-F(x)]^{n-i-1}f(x)\{i[1-F(x)]-(n-i)[F(x)]\}$$

$$=\sum_{i=k}^{n}iC_n^i[F(x)]^{i-1}[1-F(x)]^{n-i-1}f(x)-\sum_{i=k}^{n}nC_n^i[F(x)]^{i-1}[1-F(x)]^{n-i-1}[F(x)]$$

$$=\sum_{i=k}^{n}iC_n^i[F(x)]^{i-1}[1-F(x)]^{n-i-1}f(x)-\sum_{i=k}^{n-1}(i+1)C_n^{i+1}[F(x)]^i[1-F(x)]^{n-i-1}f(x)$$

$$=kC_n^k[F(x)]^{k-1}[1-F(x)]^{n-k-1}f(x)$$

$$X_{(k)}\sim\frac{n!}{(k-1)!(n-k)!}[F(x)]^{k-1}[1-F(x)]^{n-k}f(x)$$

★ 样本中位数的渐近分布（思考）：

若总体的密度函数为 $f(x)$，其中位数为 x_{med}，样本中位数记为

$$X_{med}=\begin{cases}X_{(\frac{n+1}{2})}, & n\text{为奇数}\\ \frac{1}{2}[X_{(\frac{n}{2})}+X_{(\frac{n}{2}+1)}], & n\text{为偶数}\end{cases}$$

则 $X_{med}\dot{\sim}N(x_{med},\dfrac{1}{4n\bullet f^2(x_{med})})$

样本 p 分位数 $X_p=\begin{cases}X_{([np+1])}, & np\text{为非整数}\\ \frac{1}{2}[X_{(np)}+X_{(np+1)}], & np\text{为整数}\end{cases}$

则 $X_p\dot{\sim}N(x_p,\dfrac{p(1-p)}{n\bullet f^2(x_p)})$

思考：上述渐近分布，如何利用经验分布函数（结合 Delta 方法）给出证明。

1.4 三大抽样分布：χ^2 分布、*t* 分布和 *F* 分布的概念及性质，分位数的概念

★ $U \sim \chi^2(n) \Leftrightarrow U = \sum_{i=1}^{n} X_i^2, \quad X_i (i=1,2,\cdots,n)\ i.i.d. \sim N(0,1)$

$$U \sim \chi^2(n) = Gamma(\frac{n}{2}, \frac{1}{2})$$
$$E(U) = n, D(U) = 2n$$

χ^2 分布具有可加性。

★ $T \sim t(n) \Leftrightarrow T = \frac{X}{\sqrt{\frac{Y}{n}}}, \quad X \sim N(0,1), Y \sim \chi^2(n),\ X, Y\ i.d.$

$$f_T(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1+\frac{x^2}{n})^{-\frac{n+1}{2}}, -\infty < x < +\infty$$

$n=1, \quad T \sim t(1) = C(1,0)$ 标准 **Cauchy** 分布（期望不存在），

$n>1, \ ET = 0$，

$n>2, \ DT = \frac{n}{n-2}$，

★ $F \sim F(m,n) \Leftrightarrow T = \frac{\frac{X}{m}}{\frac{Y}{n}}, \quad X \sim \chi^2(m), Y \sim \chi^2(n),\ X, Y\ i.d.$

$$f_F(x) = \frac{\Gamma(\frac{m+n}{2})(\frac{m}{n})^{\frac{m}{2}}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} x^{\frac{m}{2}-1}(1+\frac{m}{n}x)^{-\frac{m+n}{2}}, x > 0$$

$F \sim F(m,n) \Leftrightarrow \frac{1}{F} \sim F(n,m)$

★ ***X*** 的 α 分位数 ν_α：$F(\nu_\alpha) = P(X \le \nu_\alpha) = \alpha$

1.5　正态总体抽样分布定理：样本均值和样本方差的分布（单总体与双总体（主要均值差与方差比））

1) 样本均值 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \sim N(\mu,\frac{\sigma^2}{n})$，从而 $Z\equiv\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,\ 1)$

2) $K^2\equiv\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\sim\chi^2(n).$

3) $\chi^2\equiv\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$，及 $\overline{X}$ 和样本方差 S^2 的独立性.

4) $T\equiv\frac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$.

5） $\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)$；

$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}\dot{\sim}N(0,1)$（渐近分布）

6） $T_{n_1,n_2}\equiv\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2);\quad S_w^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$（方差齐性条件下）

6) $F_{n_1,n_2}=\frac{\sigma_2^2S_1^2}{\sigma_1^2S_2^2}\sim F(n_1-1,n_2-1)$.

$$\tilde{F}_{n_1,n_2}=\frac{n_2\sigma_2^2\sum_{i=1}^{n_1}(X_i-\mu_1)^2}{n_1\sigma_1^2\sum_{i=1}^{n_2}(Y_i-\mu_2)^2}\sim F(n_1,n_2)$$

主要引理：

引理：$X_1,X_2,\cdots,X_n$是来自总体$X\sim N(\mu,\sigma^2)$的简单随机样本，则$\overline{X}$与$(X_1-\bar{X},X_2-\bar{X},\cdots,X_n-\bar{X})$相互独立，（同时给出他们的联合分布）。

证明：显然$n+1$维向量$(\bar{X},X_1-\bar{X},X_2-\bar{X},\cdots,X_n-\bar{X})^T$服从 **Gauss** 分布$N(\begin{pmatrix}\mu^{(1)}\\ \mu^{(2)}\end{pmatrix},\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix})$，

其中$\mu^{(1)}=\mu,\mu^{(2)}=(0,0,\cdots,0)^T$

$$\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}=\begin{pmatrix}\frac{\sigma^2}{n} & 0 & 0 & \cdots & 0\\ 0 & \frac{(n-1)\sigma^2}{n} & -\frac{\sigma^2}{n} & \cdots & -\frac{\sigma^2}{n}\\ 0 & -\frac{\sigma^2}{n} & \frac{(n-1)\sigma^2}{n} & \cdots & -\frac{\sigma^2}{n}\\ \vdots & \vdots & \cdots & \cdots & \vdots\\ 0 & -\frac{\sigma^2}{n} & \cdots & -\frac{\sigma^2}{n} & \frac{(n-1)\sigma^2}{n}\end{pmatrix}$$

为分块对角矩阵。

事实上，

$$Cov(\bar{X},X_i-\bar{X})=Cov(\bar{X},X_i)-D\bar{X}=\frac{\sigma^2}{n}-\frac{\sigma^2}{n}=0,i=1,2,\cdots,n$$

$$\begin{aligned}&Cov(X_i-\bar{X},X_j-\bar{X})\\ &=Cov(X_i,X_j)-Cov(X_i,\bar{X})-Cov(\bar{X},X_i)+D\bar{X}\\ &=\begin{cases}\frac{(n-1)\sigma^2}{n}, & i=j\\ -\frac{\sigma^2}{n}, & i\neq j\end{cases},\end{aligned}$$

从而可知$\overline{X}$和S^2相互独立。

注：从而$\overline{X}$也与$\max\limits_{1\le i\le n}X_i-\min\limits_{1\le i\le n}X_i$相互独立。事实上，

$$\max_{1\le i\le n} X_i - \min_{1\le i\le n} X_i = \max_{1\le i\le n}(X_i - \bar{X}) - \min_{1\le i\le n}(X_i - \bar{X})$$

注（1）：由于$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(n)$与$\frac{(n-1)S^2}{\sigma^2}$相互独立，且

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

由于两个独立的χ^2分布之差（前者的自由度大于后者）仍为χ^2分布，故

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

这里用到如下结论：若$X,Y\ i.d.$，$X \sim \chi^2(m), X+Y \sim \chi^2(m+n)$，则

$Y \sim \chi^2(n)$。

证明：$\varphi_X(\theta) = \left(\frac{\frac{1}{2}}{\frac{1}{2}-i\theta}\right)^{\frac{m}{2}}$，$\varphi_{X+Y}(\theta) = \left(\frac{\frac{1}{2}}{\frac{1}{2}-i\theta}\right)^{\frac{m+n}{2}}$，由于

$\varphi_{X+Y}(\theta) = \varphi_X(\theta)\varphi_Y(\theta)$，故$\varphi_Y(\theta) = \left(\frac{\frac{1}{2}}{\frac{1}{2}-i\theta}\right)^{\frac{n}{2}}$

注（2）：若总体为对称分布（方差存在），则$\overline{X}$和S^2不相关。更一般地，

若总体的三阶中心矩$\nu_3 \triangleq E(X-EX)^3$存在，则$Cov(\bar{X}, S^2) = \frac{\nu_3}{n}$。此

外，只有正态总体时，$\overline{X}$和S^2才独立。

例 设总体$X \sim N(\mu, \sigma^2)$，则$E(S^2)$=_____ ，$D(S^2)$=_______ .

【σ^2；$\frac{2\sigma^4}{n-1}$】

例 设随机变量 $X \sim t(n)(n>1), Y=\dfrac{1}{X^2}$，则

(A) $Y \sim \chi^2(n)$ **(B)** $Y \sim \chi^2(n-1)$ **(C)** $Y \sim F(n,1)$ **(D)** $Y \sim F(1,n)$

【C】

例 设 $X_1, X_2, \cdots, X_{20}$ 是来自总体 $X \sim N(0,\sigma^2)$ 的简单随机样本，则统计量

$\sum_{i=1}^{10}(-1)^i X_i \Big/ \sqrt{\sum_{i=11}^{20} X_i^2}$ 服从的分布是 $t(10)$.

例 设总体 X 服从正态分布 $N(0,2^2)$，而 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的简单随机样本，则随机变量 $Y=\dfrac{X_1^2+\cdots+X_{10}^2}{2(X_{11}^2+\cdots+X_{15}^2)}$ 服从__________分布。 【$F(10,5)$】

例 设 $X_1, X_2, \cdots, X_{n+1}$ 是正态总体 $N(\mu, \sigma^2)$ 的简单样本，$\overline{X}=\frac{1}{n}\sum_{i=1}^{n} X_i$ 和 $S_n{}^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$

1）试求 $(n-1)(X_1-\mu)^2 \Big/ [\sum_{i=2}^{n}(X_i-\mu)^2]$ 的分布.

【$F(1,n-1)$】

2) 试求 $\dfrac{X_{n+1}-\overline{X}}{S_n}\sqrt{\dfrac{n-1}{n+1}}$ 的分布.

【$t(n-1)$】

例 设 $X_1, X_2, \cdots, X_9$，是正态总体的简单样本，令 $Y_1=\frac{1}{6}\sum_{i=1}^{6} X_i$，$Y_2=\frac{1}{3}\sum_{i=1}^{3} X_{6+i}$，$S^2=\frac{1}{2}\sum_{i=7}^{9}\left(X_i-Y_2\right)^2$ 和 $Z=\sqrt{2}\left(Y_1-Y_2\right)/S$. 试证 $Z \sim t(2)$.

第 12 周　充分统计量

二、充分统计量

2.1 定义（不损失任何关于参数θ的信息）

样本中包含的关于总体的信息可分为两部分：其一是关于总体结构的信息，即反映总体分布的结构；其二是关于总体中未知参数的信息。统计量具有压缩数据功能，一个好的统计量应该能将样本中包含未知参数的全部信息提取出来，这种不损失未知参数的信息的统计量就是我们要介绍的充分统计量。

★ 充分统计量的定义如下：设$X_1, X_2, \cdots, X_n$是总体 X 的简单样本，总体分布函数为$F(x,\theta)$，称统计量$T = T(X_1, X_2, \cdots, X_n)$为$\theta$的充分统计量，如果在给定$T$的取值后，$X_1, X_2, \cdots, X_n$的条件分布与$\theta$无关。

充分统计量说明：

- 样本X所包含的关于θ信息 = $T(X)$所包含的关于θ信息 + $T(X)$已知后X还包含的关于θ信息
- 若$T(X)$是充分统计量，则后者（$T(X)$已知后X还包含的信息）应当为零，即$T(X)$已知后，X不再包含关于θ的信息（即充分统计量的意义）。显然。全样本$T(X) = (X_1, \cdots, X_n)$是一个充分统计量，但在实际中，这个统计量的意义不大。

★ 定理 **A**：（上课略，有更好的结果）若样本$X_1, X_2, \cdots, X_n$的联合分布密度（或 **p.m.f**）为$f(x_1, x_2, ..., x_n \mid \theta)$，统计量$T = T(X_1, X_2, ..., X_n)$的分布密度（或 **p.m.f**）为$q(t \mid \theta)$，则若$\dfrac{f(x_1, x_2, ..., x_n \mid \theta)}{q(T(x_1, x_2, ..., x_n) \mid \theta)}$与$\theta$无关，则$T = T(X_1, X_2, ..., X_n)$为充分统计量。

例（离散情形）设$X_1, X_2, \cdots, X_n$是总体$X \sim B(1,\theta), 0 < \theta < 1$（即

Bernoulli 分布）的一个样本，证明$T=\sum_{i=1}^{n}X_i$为θ的充分统计量

用定义证明：只需证明$P(X_1=x_1,...,X_n=x_n|T=t)$与$\theta$无关，这里$x_i=0,1$，$t=\sum_{i=1}^{n}x_i$（否则，概率为 **0**），由于$T=\sum_{i=1}^{n}X_i\sim B(n,\theta)$

$$P(X_1=x_1,...,X_n=x_n|T=t)=\frac{P(X_1=x_1,...,X_n=x_n,T=t)}{P(T=t)}$$

$$=\frac{P(X_1=x_1,...,X_{n-1}=x_{n-1},X_n=t-\sum_{i=1}^{n-1}x_i)}{P(T=t)}$$

$$=\frac{\prod_{i=1}^{n-1}[\theta^{x_i}(1-\theta)^{1-x_i}]\bullet\theta^{t-\sum_{i=1}^{n-1}x_i}(1-\theta)^{1-(t-\sum_{i=1}^{n-1}x_i)}}{C_n^t\theta^t(1-\theta)^{n-t}}=\frac{1}{C_n^t}$$

例（连续情形）设$X_1,X_2,\cdots,X_n$是总体$X\sim N(\mu,1)$的一个样本，证明

$T=\bar{X}$为μ的充分统计量

解：利用定理 **A** 证明（利用定义也可以证明，见教材）：

$X_1,X_2,\cdots,X_n$的联合分布密度为

$$f(x_1,x_2,...,x_n|\mu)$$

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

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

$T=\bar{X}\sim N(\mu,\frac{1}{n})$的分布密度为$q(t|\mu)=\frac{1}{\sqrt{\frac{2\pi}{n}}}e^{-\frac{(t-\mu)^2}{2\bullet\frac{1}{n}}}$，故

$$\frac{f(x_1,x_2,\ldots,x_n\mid\mu)}{q(T(x_1,x_2,\ldots,x_n)\mid\mu)}$$

$$=\frac{(\frac{1}{\sqrt{2\pi}})^n e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2}}}{\frac{1}{\sqrt{\frac{2\pi}{n}}}e^{-\frac{(\bar{x}-\mu)^2}{2\cdot\frac{1}{n}}}}$$

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

与μ无关，故$T=\bar{X}$为μ的充分统计量。

例 设X_1,X_2是总体$X\sim P(\lambda)$的一个样本，证明$T_1=X_1+X_2$为λ的充分统计量，而$T_2=X_1+2X_2$不是λ的充分统计量。

解：

$$P(X_1=x_1,X_2=x_2\mid T_1=t)$$

$$=\begin{cases}\dfrac{P(X_1=x_1,X_2=t-x_1)}{P(T_1=t)}=\dfrac{\dfrac{\lambda^{x_1}e^{-\lambda}}{x_1!}\dfrac{\lambda^{t-x_1}e^{-\lambda}}{(t-x_1)!}}{\dfrac{(2\lambda)^t e^{-2\lambda}}{t!}}=\dfrac{C_t^{x_1}}{2^t}, & t=x_1+x_2,x_1,x_2=0,1,2,\cdots\\ 0, & otherwise\end{cases}$$

与λ无关，故$T_1=X_1+X_2$为λ的充分统计量。

$$P(X_1=0,X_2=1\mid T_2=2)=P(X_1=0,X_2=1\mid X_1+2X_2=2)$$

$$=\frac{P(X_1=0,X_2=1)}{P(X_1+2X_2=2)}$$

$$=\frac{P(X_1=0,X_2=1)}{P(X_1=0,X_2=1)+P(X_1=2,X_2=0)}$$

$$=\frac{e^{-\lambda}\lambda e^{-\lambda}}{e^{-\lambda}\lambda e^{-\lambda}+\frac{\lambda^2e^{-\lambda}}{2!}e^{-\lambda}}=\frac{2}{2+\lambda}$$

依赖于λ，故$T_2=X_1+2X_2$不是λ的充分统计量。

因子分解定理：
如何寻找充分统计量？（重要！！！）

★ 定理 **B**：（因子分解定理）总体概率函数 $f(x,\theta)$，样本 $(X_1,...,X_n)$，统计量 $T(X_1,...,X_n)$ 为未知参数 θ 的充分统计量的充要条件是样本在已发生情形 $(X_1,...,X_n)\to(x_1,...,x_n)$ 处的联合概率函数可以分解为 $T(x_1,...,x_n)=t$ 和 θ 的函数 $g(t,\theta)$ 与样本观察值的函数 $h(x_1,...,x_n)$ 的乘积。

即：**T** 是 **θ** 的充分统计量

$$\Leftrightarrow L(\theta)=\prod_{i=1}^{n} f(x_i;\theta)=h(x_1,x_2,...,x_n)g(T(x_1,x_2,...,x_n);\theta) \text{ 且 } \mathbf{h} \text{ 非负}$$

定理证明：只就离散情况证明（其他情形不容易证明，可略），

必要性 $\Rightarrow$：$P(X_1=x_1,...,X_n=x_n\,|\,T=t)\overset{T(x_1,...,x_n)=t}{=}h(x_1,...,x_n)$ 与 θ 无关，

由于 $\{T=t\}\supset\{X_1=x_1,\cdots,X_n=x_n\}$

因此 $P(X_1=x_1,...,X_n=x_n)=P(X_1=x_1,...,X_n=x_n\,|\,T=t)P(T=t)$

$$=h(x_1,...,x_n)P(T=t)=g(t,\theta)h(x_1,...,x_n)$$

充分性 $\Leftarrow$：现在已知 $P(X_1=x_1,...,X_n=x_n)=g(t,\theta)h(x_1,...,x_n)$，由于

$$P(T=t)=\sum_{T(x_1,\cdots,x_n)=t}P(X_1=x_1,...,X_n=x_n)$$

$$=\sum_{T(x_1,\cdots,x_n)=t}g(t,\theta)h(x_1,...,x_n)=g(t,\theta)\sum_{T(x_1,\cdots,x_n)=t}h(x_1,...,x_n)$$

因此 $P(X_1=x_1,\cdots,X_n=x_n\,|\,T=t)\overset{T(x_1,\cdots,x_n)=t}{=}\dfrac{P(X_1=x_1,\cdots,X_n=x_n,T=t)}{P(T=t)}$

$$=\frac{P(X_1=x_1,...,X_n=x_n,T=t)}{g(t,\theta)\sum\limits_{T(x_1,...,x_n)=t}h(x_1,...,x_n)}=\frac{P(X_1=x_1,...,X_n=x_n)}{g(t,\theta)\sum\limits_{T(x_1,...,x_n)=t}h(x_1,...,x_n)}$$

$$=\frac{P(X_1=x_1,...,X_n=x_n,T=t)}{g(t,\theta)\sum\limits_{T(x_1,...,x_n)=t}h(x_1,...,x_n)}=\frac{P(X_1=x_1,...,X_n=x_n)}{g(t,\theta)\sum\limits_{T(x_1,...,x_n)=t}h(x_1,...,x_n)}$$

$$=\frac{g(t,\theta)h(x_1,...,x_n)}{g(t,\theta)\sum\limits_{T(x_1,...,x_n)=t}h(x_1,...,x_n)}=\frac{h(x_1,...,x_n)}{\sum\limits_{T(x_1,...,x_n)=t}h(x_1,...,x_n)} \text{ 与 } \theta \text{ 无关。}$$

★推论：设统计量T为θ的充分统计量，统计量S与T一一对应，则S也为θ的充分统计量。

例 设$X_1,X_2,\cdots,X_n$是总体$X\sim B(1,\theta),0<\theta<1$（即 **Bernoulli** 分布），证明$T=\sum_{i=1}^{n}X_i$为$\theta$的充分统计量

解： 用因子分解证明：样本的分布为

$$P(X_1=x_1,...,X_n=x_n)=\prod_{i=1}^{n}[\theta^{x_i}(1-\theta)^{1-x_i}]=\theta^{\sum_{i=1}^{n}x_i}(1-\theta)^{n-\sum_{i=1}^{n}x_i}$$

取

$$h(x_1,x_2,...,x_n)=1,$$

$$g(T(x_1,x_2,...,x_n);\theta)=\theta^{\sum_{i=1}^{n}x_i}(1-\theta)^{n-\sum_{i=1}^{n}x_i},(T(x_1,x_2,...,x_n)=\sum_{i=1}^{n}x_i)$$

故$T=\sum_{i=1}^{n}X_i$为θ的充分统计量。

注：显然$T_1=(X_1,X_2+\cdots+X_n)$、$T_2=(X_1+X_2,X_3,X_4+\cdots+X_n)$也是$\theta$的充分统计量（不唯一）。

例 设$X_1,X_2,\cdots,X_n$是总体$X\sim N(\mu,\sigma^2)$的一个样本，μ,σ^2均未知，求(μ,σ^2)的充分统计量。

解：样本的分布密度为

$$f_{\mu,\sigma^2}(x_1,x_2,...,x_n)=\prod_{i=1}^{n}f(x_i;\mu,\sigma^2)$$

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

故$T=T(X_1,X_2,...,X_n)=(\sum_{i=1}^{n}X_i,\sum_{i=1}^{n}X_i^2)$为$(\mu,\sigma^2)$的充分统计量

注：显然$T_1=(\bar{X},S^2)$为(μ,σ^2)的充分统计量，这个一般更常用。若σ^2未

知，$\bar{X}$ 不是 μ 的充分统计量；μ 未知时，S^2 也不是 σ^2 的充分统计量。但若 σ^2 已知，$\bar{X}$ 是 μ 的充分统计量，$\mu=\mu_0$ 已知，$\sum_{i=1}^{n}(X_i-\mu_0)^2$ 是 σ^2 的充分统计量。

例 设 $X_1,X_2,\cdots,X_n$ 是总体 $X\sim U(-\frac{\theta}{2},\frac{\theta}{2})$ 的一个样本，，求参数 θ 的充分统计量。

解：样本的分布密度为

$$f(x_1,x_2,...,x_n)=\prod_{i=1}^{n}f(x_i;\theta)=\frac{1}{\theta^n}1_{\{-\frac{\theta}{2}<x_1,\cdots,x_n<\frac{\theta}{2}\}}$$

$$=\frac{1}{\theta^n}1_{\{-\frac{\theta}{2}<x_{(1)}\le x_{(n)}<\frac{\theta}{2}\}}$$

故 $(X_{(1)},X_{(n)})$ 为 θ 的充分统计量。当然全样本 $(X_1,X_2,...,X_n)$ 记顺序统计量 $(X_{(1)},X_{(2)},...,X_{(n)})$ 也是 θ 的充分统计量。

注 1：在写联合密度或分布列的时候，注意支撑，一般可用示性函数简便表示。当参数时多维的时候，充分统计量往往也是多维的，但充分统计量的维数不一定等于参数维数。

注 2：统计中还有极小充分统计量的概念，有兴趣的同学可以去了解一下（不介绍）其实，极小充分统计量是其他充分统计量的一个"加工"。

- 称 $T^*=T^*(X)$ 为极小充分统计量，若 $T^*=T^*(X)$ 为充分统计量，且对任一充分统计量 $T=T(X)$，都存在 $\varphi(\bullet)$，使 $T^*(X)=\varphi(T(X))$

三. 参数点估计

★ 问题的提出：参数估计的意义.

1 点估计

$\theta = \theta(X_1, X_2, ..., X_n)$ 点估计量

$\theta = \theta(x_1, x_2, ..., x_n)$ 点估计值

1.2. 矩估计与极大似然估计

★ 矩估计（**MOM**）：用样本矩作为总体矩估计

★ 极大似然估计（**MLE**）：使得似然函数达到极大，

$$L(x_1,\cdots,x_n;\hat{\theta}) = \max_{\theta\in\Theta} L(x_1,\cdots,x_n;\theta).$$（要注意其原理）

★ 矩估计法（**MOM**）：

1) 出发点;

★ $M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{P} \mathrm{E}X^k = \mu_k$， $k \ge 1$

2) 方法与步骤

① 求出总体的 **k** 阶原点矩: $a_k = EX^k = \int_{-\infty}^{+\infty} x^k dF(x;\theta_1,\theta_2,...,\theta_m)$

② 解方程组 $a_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ **(k=1,2,...,m)**，得 $\theta_k = \theta_k(X_1, X_2, ..., X_n)$ 即为所求

★ 极大似然估计法:

1) 出发点; 实际推断原则（似然函数达到极大）

★ 似然函数 $L(x_1,\cdots,x_n;\theta) = \begin{cases} \prod_{i=1}^{n} P(X = x_i), \\ \prod_{i=1}^{n} f(x,\theta) \end{cases}$

2) 方法与步骤

① 写出似然函数 $L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$，求出 **lnL** 及似然方程 $\left.\frac{\partial \ln L}{\partial \theta_i}\right|_{\theta=\theta} = 0$

i=1,2,...,m

② 解似然方程得到$\theta_i(x_1,x_2,...,x_n)$，即极大似然估计$\theta_i(X_1,X_2,...,X_n)$

i=1,2,...,m

注：似然方程无解时，求出θ的定义域中使得似然函数最大的值，即为最大似然估计。

极大似然估计可以不唯一（如$U[\theta-\frac{1}{2},\theta+\frac{1}{2}]$）

例 设 ***X*** ～**B(1, *p*)**. $X_1,X_2,\ldots,X_n$ 是来自 ***X*** 的一个样本，试求参数 ***p*** 的矩估计量

$\hat{p}_M$

例 设 $X \sim f(x)=\begin{cases}\frac{6x}{\theta^3}(\theta-x) & 0<x<\theta\\ 0 & else\end{cases}$，

求 **1）.** θ的矩估计量$\hat{\theta}$，**2）.** 求$\hat{\theta}$的方差。

【$2\bar{X}$; $\frac{\theta^2}{5n}$】

解：矩估计：

$$EX=\int_0^\theta x\frac{6x}{\theta^3}(\theta-x)dx=\frac{\theta}{2}\overset{令}{=}\bar{X}$$
$$\Rightarrow\hat{\theta}=2\bar{X}$$

又由于

$$EX^2=\int_0^\theta x^2\frac{6x}{\theta^3}(\theta-x)dx=\frac{3\theta^2}{10}$$
$$\Rightarrow DX=\frac{3\theta^2}{10}-(\frac{\theta}{2})^2=\frac{\theta^2}{20}$$

故 $D\hat{\theta}=D(2\bar{X})=4\frac{DX}{n}=\frac{\theta^2}{5n}$

例

设总体 ***X*** 的概率密度为 $f(x)=\begin{cases}e^{-(x-\theta)} & x\ge\theta\\ 0 & x<\theta\end{cases}$，而 $X_1,X_2,\cdots,X_n$ 是来自总体 ***X*** 的简单随机样本，求未知参数θ的矩估计量 【$\bar{X}-1$】

例 设总体 **X** 的概率分布为

$$X\sim\begin{pmatrix}0 & 1 & 2 & 3\\ \theta^2 & 2\theta(1-\theta) & \theta^2 & 1-2\theta\end{pmatrix}$$

其中$\theta(0<\theta<\frac{1}{2})$是未知参数，利用总体 **X** 的如下样本值

3，1，3，0，3，1，2，3，

求θ的矩估计值和最大似然估计值. 【**1/4;** $\frac{7-\sqrt{13}}{12}$ 】

解：矩估计：$EX=2\theta(1-\theta)+2\theta^2+3(1-2\theta)=3-4\theta\overset{令}{=}\bar{X}\quad\Rightarrow\hat{\theta}=\frac{3-\bar{X}}{4}$

由于$\bar{x}=2$ ，故θ的矩估计值为$\frac{3-\bar{x}}{4}=\frac{1}{4}$；

此时的似然函数为

$$L=P(X_1=3,X_2=1,\cdots,X_8=3)=(1-2\theta)^4[\theta^2]^1[2\theta(1-\theta)]^2[\theta^2]^1=4\theta^6(1-2\theta)^4(1-\theta)^2$$

$$\ln L=\ln 4+6\ln\theta+4\ln(1-2\theta)+2\ln(1-\theta)$$

$$\frac{\partial\ln L}{\partial\theta}=\frac{6}{\theta}-\frac{8}{1-2\theta}-\frac{2}{1-\theta}=0,\Rightarrow 12\theta^2-14\theta+3=0$$

$$\Rightarrow\hat{\theta}_{MLE}=\frac{7\pm\sqrt{13}}{12}$$

由于$0<\theta<\frac{1}{2}$，故为$\frac{7-\sqrt{13}}{12}$。

例 设总体X 的概率密度为

$$f(x;\theta)=\begin{cases}\theta, & 0<x<1,\\ 1-\theta, & 1\le x<2,\\ 0, & 其他,\end{cases}$$

其中θ是未知参数$(0<\theta<1)$，$X_1,X_2...,X_\mathrm{n}$为来自总体X 的简单随机样本，记N为样本值$x_1,x_2...,x_n$中小于 **1** 的个数，求θ的最大似然估计.

【$\hat{\theta}=\frac{N}{n}$】

例 设总体 **X** 的二阶矩存在，$X_1,X_2,\ldots,X_n$为其样本. 求 **X** 的期望μ和方差σ^2的矩估计量$\hat{\mu}_M$和$\hat{\sigma}_M^2$；又若 **X** 为正态，求最大似然估计量$\hat{\mu}_L$和$\hat{\sigma}_L^2$.

例 设总体 **X** ~$U[0,\theta]$，$\theta(>0)$未知，$X_1,X_2,\ldots,X_n$是一个样本，试求$\boldsymbol{\theta}$ 的矩估计量$\hat{\theta}_M$和最大似然估计量$\hat{\theta}_L$. 【$\hat{\theta}_M=2\bar{X}$,$\hat{\theta}_L=X_{(n)}$】

【矩估计法：易知$E(X)=\frac{\theta}{2}$，建立方程$\frac{\theta}{2}=\overline{X}\Rightarrow\hat{\theta}_M=2\overline{X}$。

最大似然估计法：$L(x_1,\cdots,x_n;\theta)=\begin{cases}\theta^{-n}, & 0\le x_i\le\theta,\quad i=1,2,\cdots,n\\ 0, & \text{其他.}\end{cases}$；由似然函数

可以看出，要使$L(x_1,\cdots,x_n;\theta)$最大，就要使θ尽可能地小，但θ又不能小于

$x_{(n)}=\max\{x_1,\cdots,x_n\}$，所以，$\theta$取$x_{(n)}=\max\{x_1,\cdots,x_n\}$时就使$L(x_1,\cdots,x_n;\theta)$达到

最大，故$\hat{\theta}_L=X_{(n)}$】

例　设总体 **X** 的概率密度函数为 $f(x)=\begin{cases}\lambda e^{-\lambda(x-\mu)} & \text{当}x>\mu\\ 0 & \text{当}x\le\mu\end{cases}$，这里$\mu$和**λ(>0)**

都是参数. 又设$X_1,X_2,\cdots,X_n$为该总体的简单样本，而$x_1,x_2,\cdots,x_n$为其样本观察值.

1) 设λ已知，求μ的极大似然估计$\hat{\mu}_L$　.　　　　　　【$\min(x_1,x_2,\cdots,x_n)$】

2) 设μ已知，求λ的矩估计$\hat{\lambda}_M$　.　　　　　　【$(\bar{x}-\mu)^{-1}$】

例　设总体的分布函数为

$$F(x;\lambda,\theta)=\begin{cases}0, & \text{当}x\le\theta,\\ 1-(\theta/x)^{\lambda}, & \text{当}\theta<x.\end{cases},$$

其中$\theta>0,\lambda>0$都是未知参数.　设$X_1,\cdots,X_n$为简单样本，求θ和λ的极大似然估计。

2. 点估计的评价标准

点估计与优良性：概念、无偏估计、均方误差准则、相合估计(一致估计)、渐近正态估计

★ 无偏性：θ的估计量$\hat{\theta}(x_1,\cdots,x_n)$满足$E\hat{\theta}<+\infty$，且$\forall\theta\in\Theta$，有$E\hat{\theta}=\theta$，称$\hat{\theta}$为$\theta$的无偏估计量。

★ θ的均方误差：$MSE(\theta,\theta)=E(\theta-\theta)^2=D\theta+(E\theta-\theta)^2=D(\hat{\theta})+(Bias(\hat{\theta}))^2$

若θ是无偏估计，则$MSE(\theta,\theta)=D\theta$

★ 有效性：对于θ的无偏估计量θ_1,θ_2，若对于任意的$\theta\in\Theta$有

$D(\theta_1) \le D(\theta_2)$，

且至少有一个$\theta \in \Theta$使得上述不等式严格成立，则称θ_1比θ_2有效。

★ 性质

☆ 设总体 ***X*** 的 ***k*** 阶矩$\mu_k = \mathrm{E}X^k,\ k \ge 1$存在. 又设$X_1, X_2, \ldots, X_n$是 ***X*** 的一个样本. 试证明不论总体服从什么分布, ***k*** 阶样本矩$M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$一定是 ***k*** 阶总体矩$\mu_k$的无偏估计. 样本方差$S^2$一定是总体方差的无偏估计。

☆ 对正态总体$N(\mu, \sigma^2)$，若 $\boldsymbol{\mu}$ 和σ^2均为未知, 则

（1）**μ** 的矩估计量和最大似然估计量$\bar{X}$是无偏的;

（2）σ^2的矩估计量和最大似然估计量$\hat{\sigma^2} = S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$是有偏的;

（3）$S^2 = \frac{n}{n-1}S_n^2$，即 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ ，是σ^2的无偏估计量.

例 设 $\boldsymbol{X} \sim U[0,\theta]$，参数$\boldsymbol{\theta}$ 未知， $X_1, X_2, \ldots, X_n$是其大小为 ***n*** 的样本. 则

1) 矩估计量 $\hat{\theta}_M = 2\bar{X}$是无偏的;

2) 似然估计$\hat{\theta}_L = X_{(n)} = \max\{X_1, X_2, \cdots, X_n\}$，不是参数$\boldsymbol{\theta}$ 的无偏估计. 但 $\frac{n+1}{n}\hat{\theta}_L = \frac{n+1}{n}X_{(n)}$是比$\hat{\theta}_M = 2\bar{X}$有效的估计量.

3）求$n(\theta - \hat{\theta}_L)$的极限分布. 【参数为$1/\theta$的指数分布】

例 设$X_1, X_2, \cdots, X_n (n > 2)$为来自总体 **N(0,**$\sigma^2$**)**的简单随机样本， $\bar{X}$ 为样本均值，记$Y_i = X_i - \bar{X}, i = 1,2,\cdots,n.$

求：（**I**） Y_i的方差$DY_i, i = 1,2,\cdots,n$；

【$\frac{n-1}{n}$】

（II）Y_1与Y_n的协方差$Cov(Y_1,Y_n)$. 【$-\frac{1}{n}$】

（III）若$c(Y_1+Y_n)^2$为σ^2的无偏估计量，求参数 ***c*.**

【$\frac{n}{2(n-2)}$】

例 设总体 ***X*** 服从 **Ex (**λ**)**，未知参数$\lambda=1/\theta$ **> 0，*pdf*** 为

$$f(x;\theta)=\begin{cases}\dfrac{1}{\theta}e^{-x/\theta}, x>0,\\ 0, \text{其它},\end{cases}$$

又设$X_1, X_2,\cdots X_n$是来自 **X** 的样本，试证

1) $\bar{X}$和$nX_{(1)}$都是θ的无偏估计量，其中$X_{(1)}=\min\{X_1, X_2,\cdots, X_n\}$.

2) 当 ***n*** **>1** 时，对于 ***θ*** 的估计，$\bar{X}$较 ***n*** $X_{(1)}$有效.

★ 相合性：$\forall\theta\in\Theta$，$\hat{\theta}_n(x_1,\cdots,x_n)\xrightarrow{P}\theta$，则称$\theta_n$为$\theta$的相合估计（一致估计）

注：相合性要求是最基本的要求，不满足相合性的估计一般不予考虑。

★ 理论依据和方法：大数定律、依概率收敛的性质

进一步的判断依据

定理 **A**：若$\lim\limits_{n\to\infty}E\theta_n=\theta$且 $\lim\limits_{n\to\infty}D\theta_n=0$，则$\theta_n$为$\theta$的相合估计。

定理 **B**: 若$\hat{\theta}_n^1,\hat{\theta}_n^2,\cdots,\hat{\theta}_n^k$分别为$\theta_1,\theta_2,\cdots,\theta_k$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$g(\hat{\theta}_n^1,\hat{\theta}_n^2,\cdots,\hat{\theta}_n^k)$为$g(\theta_1,\theta_2,\cdots,\theta_k)$的相合估计。

例 设 ***X*** ~$U[0,\theta]$，参数***θ*** 未知，$X_1, X_2,\ldots, X_n$是其容量为 ***n*** 的样本. 则其最大似然估计$\hat{\theta}_L=X_{(n)}=\max\{X_1, X_2,\cdots, X_n\}$是***θ*** 的相合估计。

解：$E(\hat{\theta}_L)=E[\max\{X_1, X_2,\cdots, X_n\}]=\dfrac{n}{n+1}\theta\to\theta$

$$D(\hat{\theta}_L)=\frac{n}{n+2}\theta^2-[\frac{n}{n+1}\theta]^2=\frac{n}{(n+1)^2(n+2)}\theta^2\to 0$$

故$\hat{\theta}_L=X_{(n)}=\max\{X_1, X_2,\cdots, X_n\}$是***θ*** 的相合估计。

第 13 周讲稿 UMVUE 和有效估计

UMVUE 和有效估计

★ θ^* 是 θ 的无偏估计，若对于 θ 的任意一个无偏估计量 θ，有 $D\theta^* \le D\theta$，则 θ^* 是 θ 的最小方差无偏估计，记 **MVUE** 或 **UMVUE**（**uniformly minimum-variance unbiased estimator**）

注：这里一致是指 $\forall\theta\in\Theta$。

★ 重要的定理（**Rao-Blackwell** 定理）

定理：随机变量 X 的方差存在，令 $\varphi(Y)=E(X\,|\,Y)$，则

$$E[\varphi(Y)]=EX\,,D[\varphi(Y)]\le DX$$

且等号成立当且仅当 $P(X=\varphi(Y))=1$。

证明：不妨设 $EX=0$，则

$$\begin{aligned}
&D[\varphi(Y)]=E[\varphi(Y)]^2\\
&=E[E(X\,|\,Y)]^2\\
&=E\{[E(X\,|\,Y)][E(X\,|\,Y)]\}\\
&=E[E(XE(X\,|\,Y)\,|\,Y)]\\
&=E(XE(X\,|\,Y))\\
&\le\sqrt{EX^2E[E(X\,|\,Y)]^2}
\end{aligned}$$

从而 $D[\varphi(Y)]=E[E(X\,|\,Y)]^2\le EX^2=DX$

注：可以直接利用 $DX=D[E(X\,|\,Y)]+E[D(X\,|\,Y)]$。

★ 推论：若 $T=T(X_1,X_2,\cdots,X_n)$ 为 θ 的充分统计量，$\hat{\theta}$ 为 θ 的任一无偏估计量，则 $\tilde{\theta}=E(\theta\,|\,T)$ 也为 θ 的无偏估计量，且 $D(\tilde{\theta})\le D(\theta)$。

例：设 $X_1,X_2,\cdots,X_n$ 是总体 $X\sim B(1,p),0<p<1$（即 **Bernoulli** 分布）的一个样本，显然估计量 X_1 是 p 的无偏估计。我们用 **Rao-Blackwell** 定理求

p 的改进的无偏估计量。

由于 $T=\sum_{i=1}^{n}X_i$ 为 p 的充分统计量，由于 X_1 只取 **0** 和 **1** 两个值，故

$$E[X_1|T=t]=P(X_1=1|T=t)$$
$$=\frac{P(X_1=1,T=t)}{P(T=t)}$$
$$=\frac{P(X_1=1,X_2+X_3+\cdots+X_n=t-1)}{P(T=t)}$$
$$=\frac{P(X_1=1)P(X_2+X_3+\cdots+X_n=t-1)}{P(T=t)}$$
$$=\frac{pC_{n-1}^{t-1}p^{t-1}(1-p)^{n-1-(t-1)}}{C_n^t p^t(1-p)^{n-t}}$$
$$=\frac{t}{n}$$

故由 **Rao-Blackwell** 定理知

$E[X_1|T]=\frac{T}{n}(=\bar{X})$ 是 p 的改进的无偏估计量。

注：利用对称性直接可知，$E[X_1|T]=E[X_1|\sum_{i=1}^{n}X_i]=\frac{\sum_{i=1}^{n}X_i}{n}$ 。

★ **UMVUE** 的一个判别准则：零无偏估计定理

如果 **UMVUE** 存在，则由 **Rao-Blackwell** 定理的推论知，它一定是充分统计量的
函数。

零无偏估计量是指期望为 **0** 的统计量。

定理（零无偏估计定理）如果 $\hat{\theta}$ 是 θ 的方差有限的无偏估计量，且对任何零无偏估计量 $\varphi=\varphi(X_1,\cdots,X_n)$，都有

$$Cov_\theta(\hat{\theta},\varphi)=0,\forall\theta\in\Theta$$

则 $\hat{\theta}$ 是 θ 的 **UMVUE**。

证明：对θ的任意一个无偏估计量$\tilde{\theta}$，显然$\varphi \triangleq \tilde{\theta}-\hat{\theta}$为零无偏估计量，且

$$D(\tilde{\theta})=D(\hat{\theta}+\varphi)=D(\hat{\theta})+D(\varphi)+2Cov(\hat{\theta},\varphi)$$
$$=D(\hat{\theta})+D(\varphi)\geq D(\hat{\theta}),\forall\theta\in\Theta$$

例：设 $X\sim E(\frac{1}{\theta})$，参数$\boldsymbol{\theta}$ 未知，$X_1,X_2,\ldots,X_n$是其容量为 **n** 的样本. 则$\bar{X}$为θ的 **UMVUE**。

解：由因子分解定理，$T=T(X_1,X_2,...,X_n)=\sum_{i=1}^{n}X_i$为$\theta$的充分统计量，

$\bar{X}=\frac{T}{n}$为θ的无偏估计量。设任何零无偏估计量$\varphi=\varphi(X_1,\cdots,X_n)$，由

于$E\varphi=\int_0^{\infty}\cdots\int_0^{\infty}\varphi(x_1,\cdots,x_n)\prod_{i=1}^{n}[\frac{e^{-\frac{x_i}{\theta}}}{\theta}]dx_1\cdots dx_n=0$

即$\int_0^{\infty}\cdots\int_0^{\infty}\varphi(x_1,\cdots,x_n)e^{-\sum_{i=1}^{n}\frac{x_i}{\theta}}dx_1\cdots dx_n=0$

两边对θ求导得

$$\int_0^{\infty}\cdots\int_0^{\infty}\frac{n\bar{x}}{\theta^2}\varphi(x_1,\cdots,x_n)e^{-\sum_{i=1}^{n}\frac{x_i}{\theta}}dx_1\cdots dx_n=0$$

即$E(\bar{X}\varphi)=0$，即$Cov(\bar{X},\varphi)=0$，故$\bar{X}$为θ的 **UMVUE**。

Cremer-Rao 不等式和有效估计

★ **Fisher 信息量：**

设总体的密度函数（或 **pmf**） $f(x;\theta)$， $\theta\in\Theta$ 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 $\dfrac{\partial}{\partial\theta}f(x;\theta)$ 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，即

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

5) 期望 $I(\theta)=E\left[\dfrac{\partial}{\partial\theta}\ln f(X;\theta)\right]^2$ 存在。

则称该期望 $I(\theta)$ 为总体分布的 **Fisher** 信息量。

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

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

★ **Cremer-Rao** 不等式

设 $X_1,X_2,\cdots,X_n$ 为取自具有 **pdf**（或 **pmf**） $f(x;\theta)$， $\theta\in\Theta=\{\theta:a<\theta<b\}$ 的总体 X 的一个样本， a， b 为已知常数， a 可以取 $-\infty$， b 可以取 $+\infty$。又 $\eta=\eta(X_1,X_2,\cdots,X_n)$ 是 $g(\theta)$ 的一个无偏估计，且满足正则条件：

1) 集合 $\{x:f(x;\theta)>0\}$ 与 θ 无关；

2) $g'(\theta)$ 与 $\dfrac{\partial f(x;\theta)}{\partial\theta}$ 存在，且对一切 $\theta\in\Theta$，

$$\frac{\partial}{\partial\theta}\int f(x;\theta)dx=\int\frac{\partial f(x;\theta)}{\partial\theta}dx,$$

$$\frac{\partial}{\partial\theta}\int\int\cdots\int\eta(x_1,x_2,\cdots,x_n)f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta)dx_1dx_2\cdots dx_n$$

$$=\int\int\cdots\int\eta(x_1,x_2,\cdots,x_n)\frac{\partial}{\partial\theta}\left[\prod_{i=1}^{n}f(x_i;\theta)\right]dx_1dx_2\cdots dx_n;$$

3) 令$I(\theta)=E_\theta\left(\frac{\partial \ln f(X;\theta)}{\partial\theta}\right)^2>0$，

则 $D_\theta\eta\ge\frac{[g'(\theta)]^2}{nI(\theta)}$ （*）

并且存在一个有可能依赖于θ但不依赖于$X_1,X_2,\cdots,X_n$的数K，使得等式

$\sum_{i=1}^{n}\frac{\partial \ln f(X_i;\theta)}{\partial\theta}=K(\eta-g(\theta))$以概率 **1** 成立，以上这个条件为式（*）中等式成立的充要条件。特别地当$g(\theta)=\theta$时，不等式（*）化为

$$D_\theta\eta\ge\frac{1}{nI(\theta)}.$$

[注]1.若ξ是离散型随机变量，$f(x;\theta)$则表示为$P(\xi=x;\theta)$，相应的积分号改为求和号。

2.在使用 **R-C** 不等式时可不必验证 **2**）是否成立，因为在一般情况下，当 **1**）成立时 **2**）自动满足。

3. 满足 **1**）、**2**）假定的估计量称为正则估计。

4. R-C 下界$\frac{[g'(\theta)]^2}{nI(\theta)}$不是所有无偏估计的下界，而是无偏估计类中一个子集——正则无偏估计的方差下界。

5. R-C 的重要作用—达到 **R-C** 下界的估计量一定是 **UMVUE.**反之不然。即 **UMVUE** 不一定达到 **R-C** 下界。

6. $I(\theta)$信息量的意义

当 $D\eta=\frac{1}{nI(\theta)}$ $I(\theta)$越大，$D(\eta)$越小估计精度高，而$I(\theta)$大，则认为模型本身所含的信息量较多，或者说θ易认识，所以可视$I(\theta)$反应了模型中含有信息的量。

7.若 **C-R** 不等式的等号成立（达到 **C-R** 下界），则称$\eta=\eta(X_1,X_2,\cdots,X_n)$是$g(\theta)$的有效估计，有效估计一定是 **UMVUE**。

8. 设T_1,T_2为θ的两个无偏估计，其方差均存在，称

$$eff_\theta(T_1|T_2) \triangleq \frac{D_\theta(T_2)}{D_\theta(T_1)}$$

为T_1关于T_2的效率，称T_1比T_2有效，若$eff_\theta(T_1|T_2)>1$。

若T为θ的有效估计（即$D_\theta(T)=\dfrac{1}{nI(\theta)}$），则定义$\theta$的无偏估计$T_1$的效率为

$$eff_\theta(T_1)=eff_\theta(T_1|T)=\frac{D_\theta(T)}{D_\theta(T_1)}=\frac{1}{nI(\theta)D_\theta(T_1)}$$

故有效估计的效率为 **1**，任一无偏估计的效率不超过 **1.**

例 设样本$X_1,X_2,...,X_n$是来自$B(1,p)$，求p的 **UMVUE**。

解：设总体分布为$f(x;p)=p^x(1-p)^{1-x}$，$x=0,1,0<p<1$.

易证$\{f(x;p)|p\in p(0,1)\}$满足正则条件，因为

$$I(p)=E_p\left[\frac{\partial}{\partial p}\ln f(X,p)\right]^2=E_p\left[\frac{X-p}{p(1-p)}\right]^2=\frac{1}{p^2(1-p)^2}E_p(X-p)^2=\frac{1}{p(1-p)},$$

故$\dfrac{1}{nI(p)}=\dfrac{p(1-p)}{n}$为$p$的无偏估计的 **C-R** 下界；

$\bar{X}$作为p的无偏估计，有：$D_p(\bar{X})=\dfrac{1}{n^2}\sum_{i=1}^n D_p(X_i)=\dfrac{1}{n^2}np(1-p)=\dfrac{p(1-p)}{n}$,

$\bar{X}$达到了 **C-R** 下界，故$\bar{X}$为p的 **UMVUE**。

例 设总体X的密度函数为$f(x;\lambda)=\begin{cases}\lambda e^{-\lambda x}, & x>0;\\ 0, & x\le 0;\end{cases}$未知参数$\lambda>0$，总体的一个样本为$(X_1,X_2,\cdots,X_n)$，证明$\bar{X}$是$\dfrac{1}{\lambda}$的一个 **UMVUE**。

证明：由指数分布的总体满足正则条件可得：

$$I(\lambda)=-E\left[\frac{\partial^2}{\partial\lambda^2}\ln f(X;\lambda)\right]=-E\left(\frac{-1}{\lambda^2}\right)=\frac{1}{\lambda^2},$$

$\dfrac{\left[\left(\dfrac{1}{\lambda}\right)'\right]^2}{nI(\lambda)}=\dfrac{\left[\dfrac{-1}{\lambda^2}\right]^2}{n\dfrac{1}{\lambda^2}}=\dfrac{1}{n\lambda^2}$为$\dfrac{1}{\lambda}$的无偏估计方差的 **C-R** 下界；

$$\sqrt{n}\left(\hat{\theta}_{MLE}-\theta\right) \xrightarrow{D} N\left(0, \frac{1}{I_n(\theta)}\right)$$

另一方面$E\left(\bar{X}\right)=\dfrac{1}{\lambda}$，$V_{ar}\left(\bar{X}\right)=\dfrac{1}{n\lambda^2}$；即 $\bar{X}$ 的方差达到了 **C-R** 下界，

故 $\bar{X}$ 是$\dfrac{1}{\lambda}$的一个 **UMVUE**。

完备统计量与 **Lemann-Scheffe** 定理

定义：称分布族$\{f_\theta(x):\theta\in\Theta\}$是完备的，若$\forall\theta\in\Theta$,

$$E_\theta(g(X))=0 \Rightarrow P_\theta(g(X)=0)=1$$

定义：称统计量$T=T(X_1,X_2,\cdots,X_n)$为完备的，若T 的分布族是完备的。

注：完备分布族条件下，$E_\theta(g_1(X))=E_\theta(g_2(X))$，则有

$$P_\theta(g_1(X)=g_2(X))=1$$

例：二项分布族$\{B(n,p):0<p<1\}$（n已知）是完备分布族。

证明：若$E_p(g(X))=\sum_{k=0}^{n}g(k)C_n^k p^k(1-p)^{n-k}=0$，　$0<p<1$

则$\sum_{k=0}^{n}g(k)C_n^k(\dfrac{p}{1-p})^k=0$，　$0<p<1$

等式左面是$\dfrac{p}{1-p}$的n次多项式，故$g(k)=0,k=0,1,2,\cdots,n$，证毕。

★ $\prod_{i=1}^{n}f(x_i;\theta)=C(\theta)h(x_1,x_2,...,x_n)\exp\{b(\theta)T(x_1,x_2,...,x_n)\}$

则$T=T(x_1,x_2,...,x_n)$是θ的充分完备统计量

★ $\prod_{i=1}^{n}f(x_i;\theta)=C(\theta)h(x_1,x_2,...,x_n)\exp\{b_1(\theta)T_1(x_1,x_2,...,x_n)+b_2(\theta)T_2(x_1,x_2,...,x_n)\}$

则(T_1,T_2)是$\theta=(\theta_1,\theta_2)$的充分完备统计量

★ **Lehmann-Scheffe** 定理：

若 T 是θ的充分完备统计量，$\varphi(X_1,\cdots,X_n)$是$g(\theta)$的一个无偏估计，则

$E[\varphi(X_1,\cdots,X_n)|T]$为$g(\theta)$的惟一的 **UMVUE**

指数族分布

定义：$f(x,\theta)=h(x)e^{\theta^T T(x)-b(\theta)}$，$\theta\in\Theta$

Θ有内点.

则$(T_1(x),\cdots,T_k(x))$为完备充分统计量

★ **UMVUE** 的求解步骤：

① 求出参数θ的充分完备统计量 **T**

② 求出 $ET = g(\theta)$，则 $\theta = g^{-1}(T)$ 是 θ 的一个无偏估计

或求出一个无偏估计，然后改写成用 **T** 表示的函数

③ 综合，$E[g^{-1}(T)|T] = g^{-1}(T)$ 是 θ 的 **UMVUE**

或者：求出 θ 的矩估计或 ML 估计，再求效率，为 1 则必为 UMVUE

例：设样本 $X_1, X_2, ..., X_n$ 是来自 $B(1, p)$，求 p 的 **UMVUE**

解：（1）由上例知，$T = \sum_{i=1}^{n} X_i \sim B(n, p)$ 为充分完备统计量（也是指数族分布），且 X_1 为 p 的无偏估计。由 **Lehmann-Scheffe** 定理，p 的 **UMVUE** 为

因子分解　指数族分布

$$E(X_1 | T) = E(X_1 | \sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

例：设样本 $X_1, X_2, ..., X_n$ 是来自 $N(\theta, 1)$，求 θ 的 **UMVUE**

解：$T = \bar{X} \sim N(\theta, 1)$ 是充分完备统计量（指数族分布），且 X_1 为 θ 的无偏估计。由 **Lehmann-Scheffe** 定理，θ 的 **UMVUE** 为 $E(X_1 | T)$。

由于 $\begin{pmatrix} X_1 \\ \bar{X} \end{pmatrix} \sim N(\begin{pmatrix} \theta \\ \theta \end{pmatrix}, \begin{pmatrix} 1 & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} \end{pmatrix})$，故

$$E(X_1 | T) = E(X_1 | \bar{X}) = \theta + \frac{Cov(X_1, \bar{X})}{D\bar{X}}(\bar{X} - \theta)$$
$$= \theta + (\bar{X} - \theta) = \bar{X}$$

即 $\bar{X}$ 为 θ 的 **UMVUE**。

第 14 周讲稿（1） Bayes 估计

在统计学中有两个大的学派：频率学派（经典学派）和 **Bayes** 学派

1. 统计推断的基础

经典统计使用到两种信息：总体信息和样本信息；
Bayes 统计认为，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。

Bayes 统计的基本观点是：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 **Bayes** 公式结合起来得到一个关于未知量θ新的分布——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。

2. Bayes 公式的密度函数形式

（**1**）总体依赖于参数θ的概率函数在经典统计中记为 $f(x;\theta)$，它表示参数空间 **Θ** 中不同的θ对应不同的分布**.**在 **Bayes** 统计中应记为 $f(x\,|\,\theta)$，它表示在 **θ** 取某个给随机变量定值时总体的条件概率函数**.**

（**2**）根据参数 **θ** 的先验信息确定先验分布 $\pi(\theta)$ **.**

（**3**）从 **Bayes** 观点看，样本 $X=(x_1,\cdots,x_n)$ 的产生要分两步进行，

首先设想从先验分布 $\pi(\theta)$ 产生一个样本 θ_0；

其次从 $f(x\,|\,\theta_0)$ 中产生一组样本**.**这时样本 $X=(x_1,\cdots,x_n)$ 的联合条件概率函数为

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

这个分布综合了总体信息和样本信息**.**

（**4**）由于 θ_0 是设想出来的，仍然是未知的，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 和参数 θ 的联合分布为

$$h(X,\theta)=\pi(\theta\,|\,X)m(X)$$

这个联合分布把总体信息、样本信息和先验信息三张可用信息都综合进去了。

（**5**）我们的目的是要对未知参数θ作统计推断。在没有样本信息时，我们只能依据先验分布对θ作出推断。在有了样本观察值 $X=(x_1,\cdots,x_n)$ 之后，我们应依据 $h(X,\theta)$ 对 θ 作出推断。若把 $h(X,\theta)$ 作如下分解：

$$h(X,\theta)=\pi\left(\theta\,|\,X\right)m(X)$$

其中$m(X)$是X的边际概率函数，即

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

它与θ无关，或者说$m(X)$中不含θ的任何信息**.**因此能用来对θ作出推断的仅是条件分布$\pi(\theta\,|\,X)$，它的计算公式是

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

$$\pi(\theta\,|\,X)\propto f(X\,|\,\theta)\pi(\theta)$$

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，用总体和样本对先验分布$\pi(\theta)$作调整的结果，它要比$\pi(\theta)$更接近θ的实际情况**.**

3. Bayes 估计

由后验分布$\pi(\theta\,|\,X)$估计θ有三种常用的方法：

- 使用后验分布的密度函数最大值做为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**.**

用的最多的是后验期望估计，它一般也简称为 **Bayes** 估计，记为$\hat{\theta}_B$ **.**

例 **1** 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

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

假若我们在试验前对事件 ***A*** 没有什么了解，就使用均匀分布$U(0,1)$作为θ的先验分布，先写出 ***X*** 和θ的联合分布$h(X,\theta)=\pi\left(\theta\,|\,X\right)m(X)$

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

$$\pi(\theta\,|\,X)\propto\theta^{(x+1)-1}(1-\theta)^{(n-x+1)-1},\ \ 0<\theta<1.$$

$$\theta \mid x \sim Beta(x+1, n-x+1)$$

详细过程：X 的边际概率函数

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

故θ的后验分布为

$$\pi(\theta \mid X) = \frac{h(X,\theta)}{m(X)} = \frac{\Gamma(x+1)\Gamma(n-x+1)}{\Gamma(n+2)}\ \theta^{(x+1)-1}(1-\theta)^{(n-x+1)-1},\ \ \mathbf{0}<\theta<\mathbf{1}.$$

由于$\theta \mid x \sim Be(x+1, n-x+1)$，其后验期望估计为

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

注：假如不用先验信息，只用总体信息和样本信息，那么事件 **A** 发生的概率的最大似然估计为

$$\hat{\theta}_M = \frac{x}{n},$$

它与 **Bayes** 估计是不同的两个估计**.**某些场合，**Bayes** 估计要比最大似然估计更合理一点**.**

例 **2** 设$x_1, \cdots, x_n$是来自正态分布$N\left(\mu, \sigma_0^2\right)$的一个样本，其中$\sigma_0^2$已知，$\mu$未知，假设$\mu$的先验分布亦为$N\left(\theta, \tau^2\right)$，其中先验均值$\theta$和先验方差$\tau^2$均已知，试求$\mu$的 **Bayes** 估计。

解：样本 X 的分布和 μ 的先验分布分别为

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

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

由此可以写出 X 与 μ 的联合分布

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

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

$$\begin{aligned} h(X,\mu)&=k_1\exp\left\{-\frac{1}{2}\left[A\mu^2-2B\mu+C\right]\right\} \\ &=k_1\exp\left\{-\frac{(\mu-B/A)^2}{2/A}-\frac{1}{2}\left(C-\frac{B^2}{A}\right)\right\} \end{aligned}$$

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

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

应用 **Bayes** 公式即可得到后验分布

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

这说明在样本给定后， μ的后验分布为$N(B/A,1/A)$，即

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

后验均值即为其 **Bayes** 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。当总体方差σ_0^2较小或样本量n较大时，样本均值$\bar{x}$的权重较大；当先验方差τ^2较小时，先验均值θ的权重较大，这一综合很符合人们的经验，也是可以接受的。

例 随机变量$X\sim Ge(p)$,参数p的先验分布为$U(0,1)$

（**1**）若只对X做一次观察，其观测值为 **3**，求p的后验期望估计；

（**2**）若对X做三次观察，其观测值为 **3**，**2**，**5**，求p的后验期望估计。

解：由题意可知 $\pi\left(p\right)=1, 0<p<1$

设 $X_1, X_2, ..., X_n$ **是从随机变量 X 中抽取的随机样本，则**

$$f(x_1,\cdots,x_n|p)=\prod_{i=1}^{n}P(X=x_i|p)=\prod_{i=1}^{n}[(1-p)^{x_i-1}p]$$
$$=p^n(1-p)^{\sum_{i=1}^{n}x_i-n}, x_i=1,2,\cdots$$

从而有 $\pi\left(p|x\right)\propto f(x|p)\bullet\pi\left(p\right)\propto p^n(1-p)^{\sum_{i=1}^{n}x_i-n}, 0<p<1$**，这里**

$x=(x_1,x_2,...,x_n)$。

所以 $p|x\sim Beta(n+1,\sum_{i=1}^{n}x_i-n+1)$

① **由题意可知 n=1,x=3**

$\therefore\ p|x\sim Beta(2,3)$

$\therefore\ \hat{p}_{BE}=\frac{2}{2+3}=\frac{2}{5}$

② **由题意可知** $n=3, x_1=3, x_2=2, x_3=5$

$\therefore\ p|x\sim Beta(4,8)$

$\therefore\ \hat{p}_{BE}=\frac{4}{4+8}=\frac{1}{3}$

4 共轭先验分布

定义：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** $\pi(\theta|X)$ **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）。**

在例 1 中，由于 $U(0,1)=Beta(1,1)$**，其对应的后验分布则是** $Beta\left(x+1,n-x+1\right)$**。更一般地，设** θ **的先验分布是** $Be(a,b), a>0, b>0$**，** a**，** b **均已知，则由 Bayes 公式可以求出后验分布为** $Be\left(x+a,n-x+b\right)$**，这说明 Beta 分布是伯努利试验中成功概率的共轭先验分布。**

在例 2 中，在方差已知时正态总体均值的共轭先验分布是正态分布。

共轭先验分布列表

总体分布 $f(x\|\theta)$	先验分布 $\pi(\theta)$	后验分布 $\pi(\theta\|x)$
$N(\mu,\sigma^2)$，σ^2 已知。$\theta=\mu$	$\pi(\mu)\leftrightarrow N(\alpha,\tau^2)$	$N(\frac{n\bar{x}\sigma^{-2}+\alpha\tau^{-2}}{n\sigma^{-2}+\tau^{-2}},\frac{1}{n\sigma^{-2}+\tau^{-2}})$
$\Gamma(\alpha,\beta)$，α 已知，$\theta=\beta$	$\pi(\beta)\leftrightarrow\Gamma(\alpha_0,\beta_0)$	$\Gamma(\alpha_0+\alpha n,\beta_0+\sum_{i=1}^{n}x_i)$
$B(n_0,p)$；$\theta=p$	$\pi(p)\leftrightarrow Beta(a,b)$	$Beta(a+\sum_{i=1}^{n}x_i,b+nn_0-\sum_{i=1}^{n}x_i)$
$P(\lambda)$；$\theta=\lambda$	$\pi(\lambda)\leftrightarrow\Gamma(\alpha,\beta)$	$\Gamma(\alpha+\sum_{i=1}^{n}x_i,\beta+n)$
$NB(r,p)$；$\theta=p$	$\pi(p)\leftrightarrow Beta(a,b)$	$Beta(a+nr,b+\sum_{i=1}^{n}x_i-nr)$
$N(\mu,\sigma^2)$，μ 已知；$\theta=\sigma^2$	$\pi(\sigma^2)\leftrightarrow IGa(\alpha,\lambda)$	$IGa(\alpha+\frac{n}{2},\lambda+\sum_{i=1}^{n}(x_i-\mu)^2)$

$X\sim IGa(\alpha,\lambda)$（逆 **Gamma** 分布：

$\xi\sim\Gamma(\alpha,\lambda)\Leftrightarrow\xi^{-1}\sim IGa(\alpha,\lambda)$），如果它的密度函数为

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

区间估计

参数估计分为点估计和区间估计，前面我们已经对众多点估计方法有了比较详细的介绍。

点估计是用一个统计量来作为参数的估计，区间估计是找两个统计量 $\hat{\theta}_L=\hat{\theta}_L(X_1,X_2,\cdots,X_n)$ 和 $\hat{\theta}_U=\hat{\theta}_U(X_1,X_2,\cdots,X_n)$，其中对任何样本观测值都有 $\hat{\theta}_L\le\hat{\theta}_U$，并用区间 $[\hat{\theta}_L,\hat{\theta}_U]$ 作为参数的区间估计（量）。（注意，闭区间不是必要的，可以是开区间也可以是左开右闭等等）

例　$X_1,X_2,\cdots,X_9$ 是取自正态总体 $N(\mu,1)$ 的一个样本，那么 $[\bar{X}-1,\bar{X}+1]$ 就是参数 μ 的一个区间估计，在点估计部分，我们用样本均值 $\bar{X}$ 作为 μ 的点估计量，在很多场合下，我们需要得到 μ 的一个（随机）区间估计。

$$
\begin{aligned}
&P(\mu\in[\bar{X}-1,\bar{X}+1])=P(-1\le\bar{X}-\mu\le 1)\\
&=P\left(-3\le\frac{\bar{X}-\mu}{\sqrt{\frac{1}{9}}}\le 3\right)\\
&=2\Phi(3)-1\approx 0.997
\end{aligned}
$$

也就是说，随机区间 $[\bar{X}-1,\bar{X}+1]$ 涵盖参数真值的概率是 **0.997.**

区间估计（置信区间）

★　区间估计的概念与求法

随机区间涵盖参数真值的概率称为置信度，即下面这个概率

$$P_\theta(\theta\in[\hat{\theta}_L(X_1,X_2,\cdots,X_n),\hat{\theta}_U(X_1,X_2,\cdots,X_n)])$$

这个概率一般来说与 θ 有关，希望最小的置信度（置信水平或系数）也比较大，即 $\inf\limits_{\theta\in\Theta}P_\theta(\theta\in[\hat{\theta}_L,\hat{\theta}_U])$ 比较大。

除了置信系数的要求，当然我们希望置信区间的平均长度$E_\theta(\hat{\theta}_U-\hat{\theta}_L)$越小越好。即置信水平给定时，置信区间的平均长度越小越好，即置信度越高。

Neyman 准则：在保证置信水平的前提下，最求高的置信度。

★ 概念的引出和理论依据.

- 置信水平为$1-\alpha$（双侧）置信区间：

$$P_\theta(\hat{\theta}_L(X_1,\cdots,X_n)<\theta<\hat{\theta}_U(X_1,\cdots,X_n))\ge 1-\alpha,\forall\theta\in\Theta$$

同等置信区间：

$$P_\theta(\hat{\theta}_L(X_1,\cdots,X_n)<\theta<\hat{\theta}_U(X_1,\cdots,X_n))=1-\alpha,\forall\theta\in\Theta$$

- 置信水平为$1-\alpha$单侧置信区间：

单侧置信上限：$P_\theta(\theta<\hat{\theta}_U(X_1,\cdots,X_n))\ge 1-\alpha,\forall\theta\in\Theta$

单侧置信下限：$P_\theta(\theta>\hat{\theta}_L(X_1,\cdots,X_n))\ge 1-\alpha,\forall\theta\in\Theta$

同等单侧置信限区间：改为等号即可

★ 置信区间的一般求法（枢轴变量法）：

（**1**） 先找一个与要估计的参数θ（或$g(\theta)$）有关的统计量 ***T*** ，一般是一良好的点估计
（**MLE** 或充分统计量）；

（**2**） 寻找一样本$X_1,X_2,\cdots,X_n$的函数，即找出 ***T*** 和θ（或$g(\theta)$）的某一函数：

$$Z=Z(X_1,X_2,\cdots,X_n,\theta)\equiv Z(T,\theta)$$

它只含待估参数和样本，不含其它未知参数，并且 **Z** 的分布（或渐近分布）**G** 已知，且不依赖于任何未知参数（也不依赖于待估参数θ）**(**称 ***Z*** 为枢轴变量**)**;

（**3**） 对于给定的置信度$1-\alpha$，定出两常数 ***a***，***b***，使得

$$P(a\le Z(X_1,X_2,\cdots,X_n,\theta)\le b)=1-\alpha$$

a，***b*** 选用原则：

*****：$E_\theta(\hat{\theta}_U-\hat{\theta}_L)$最小（最优置信区间）

*：一般 ***a***，***b*** 选用分布（或渐近分布）G 的 $\frac{\alpha}{2}$ 和 $1-\frac{\alpha}{2}$ 分位点（等尾置信区间）；

（4） 若能从 $a \le Z(X_1, X_2, \cdots, X_n, \theta) \le b$，得到等价的不等式

$$\hat{\theta}_L(X_1, X_2, \cdots, X_n) \le \theta \le \hat{\theta}_U(X_1, X_2, \cdots, X_n)$$

$$(\text{或}\,\tilde{\theta}_L(X_1, X_2, \cdots, X_n) \le g(\theta) \le \tilde{\theta}_U(X_1, X_2, \cdots, X_n))$$

那么 $[\hat{\theta}_L, \hat{\theta}_U]$ (或 $[\tilde{\theta}_L, \tilde{\theta}_U]$) 就是 θ（或 $g(\theta)$）的一个置信度为 $1-\alpha$ 的置信区间.

注：枢轴变量的分布为单峰对称分布时，最优置信区间=等尾置信区间

例：设样本 $X_1, \cdots, X_8$ 是来自 $U[0,\theta]$，求 θ 的置信度为 $(1-\alpha)\%$ 的等尾置信区间与最优置信区间。

解：

Step 1：从充分统计量或点估计出发找枢轴量

$X_{(n)}$ 为 θ 的充分统计量（是 **MLE**）

$\hat{\theta}_{MLE} = \bar{X}_n \sim (\frac{x}{\theta})^n, 0<x<\theta$

枢轴量为 $Y = \frac{X_{(n)}}{\theta} \sim f_Y(y) = 8y^7 1_{\{0 \le y \le 1\}}$

即 $0 < \frac{x}{\theta} < 1$

Step 1：确定常数

等尾置信区间：

$$P(y_{\frac{\alpha}{2}} \le \frac{X_{(n)}}{\theta} \le y_{1-\frac{\alpha}{2}}) = 1-\alpha$$

其中 $\int_0^{y_{\frac{\alpha}{2}}} 8y^7 dy = \frac{\alpha}{2}$，$y_{\frac{\alpha}{2}} = (\frac{\alpha}{2})^{\frac{1}{8}}$，同理 $y_{1-\frac{\alpha}{2}} = (1-\frac{\alpha}{2})^{\frac{1}{8}}$

Step 3：不等式改写

等尾置信区间为 $[\frac{X_{(n)}}{(1-\frac{\alpha}{2})^{\frac{1}{8}}}, \frac{X_{(n)}}{(\frac{\alpha}{2})^{\frac{1}{8}}}]$

同上述步骤：最优置信区间：

$$P(a \le \frac{X_{(n)}}{\theta} \le b) = 1-\alpha$$
$$b^8 - a^8 = 1-\alpha$$
$$\min\ (\frac{1}{a}-\frac{1}{b})EX_{(n)}$$
$$s.t.\quad b^8 - a^8 = 1-\alpha$$
$$\quad\quad 0 \le a < b \le 1$$

求条件极值问题（可忽略$EX_{(n)}$），解得$b=1, a=\sqrt[n]{\alpha}\quad (n=8)$

即最优置信区间：$[X_{(n)}, \frac{X_{(n)}}{\sqrt[n]{\alpha}}]$。

例：设样本$X_1,\cdots,X_n$是来自$E(\frac{1}{\theta})$，求θ的置信度为$(1-\alpha)\%$的同等置信下限。

解：**Step 1**：从充分统计量或点估计出发找枢轴量

由于$T=\sum_{i=1}^{n}X_i$是θ的充分统计量，$T=\sum_{i=1}^{n}X_i \sim \Gamma(n,\frac{1}{\theta})$

枢轴量为$\frac{2T}{\theta} \sim \Gamma(n,\frac{1}{2}) = \chi^2(2n)$（非对称分布）

Step 1：确定常数

$$P(\frac{2T}{\theta} \le \chi^2{}_{1-\alpha}(2n)) = 1-\alpha$$

Step 3：不等式改写

$$P(\theta \ge \frac{2T}{\chi^2{}_{1-\alpha}(2n)}) = 1-\alpha$$

故θ的置信度为$(1-\alpha)\%$的等尾置信下限为$\hat{\theta}_L = \frac{2T}{\chi^2{}_{1-\alpha}(2n)}$

★〖等尾双侧置信区间与单侧置信区间的对照〗 $\alpha \leftrightarrow \frac{\alpha}{2}$

例：设样本$X_1,\cdots,X_n$（大样本）是来自$B(1,p)$，求p的置信度为$(1-\alpha)\%$的等尾置信区间。

解：**Step 1**：从充分统计量或点估计出发找枢轴量

由于$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$是$\theta$的充分统计量，

为简化问题，找枢轴量需要用到中心极限定理：$\frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}}\sim AN(0,1)$ （渐近正态）

枢轴量为$\frac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}}\sim AN(0,1)$

Step 2：确定常数

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

Step 3：不等式改写

$$p\in\frac{1}{1+\frac{\lambda}{n}}[\bar{X}+\frac{\lambda}{2n}\pm\sqrt{\frac{\bar{X}(1-\bar{X})}{n}\lambda+(\frac{\lambda}{2n})^2}]\qquad(\lambda=u_{1-\frac{\alpha}{2}}^2)$$

由于n很大，故θ的置信度为$(1-\alpha)\%$的等尾置信区间为

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

★ 〖等尾双侧置信区间与单侧置信区间的对照〗 $\alpha\leftrightarrow\frac{\alpha}{2}$

★ 正态总体参数的置信区间

1)求μ的置信度为 1-α的置信区间.

（a）设方差σ^2已知；(b) 设方差σ^2未知

2) 方差σ^2的置信度$1-\alpha$置信区间

(a) 设μ已知　　(b) 设μ未知

3） 两个总体均值差$\mu_1-\mu_2$的置信区间

(a) 设$\sigma_1{}^2,\sigma_2{}^2$均为已知

(b) 设$\sigma_1{}^2=\sigma_2{}^2=\sigma^2$，但$\sigma^2$为未知.

(c) 设$\sigma_1{}^2,\sigma_2{}^2$均为未知，且不知它们是否相等

4) 两个总体方差比$\sigma_1{}^2/\sigma_2{}^2$的置信区间

总体均值的区间估计	条件	枢轴量	双侧等尾置信区间
	1、正态总体，方差已知	$Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$	$\left(\bar{X}-u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X}+u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$
	2、正态总体，方差未知	$T=\frac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$	$\left(\bar{X}-t_{1-\alpha/2}\frac{S}{\sqrt{n}},\ \bar{X}+t_{1-\alpha/2}\frac{S}{\sqrt{n}}\right)$
	3、双正态总体均值差，σ_1^2,σ_2^2均已知	$Z=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$ $\sim N(0,1)$	$(\bar{X}_1-\bar{X}_2)\pm u_{1-\alpha/2}\cdot\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$
	4、双正态总体均值差，$\sigma_1^2=\sigma_2^2$未知	$T=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ $\sim t(n_1+n_2-2)$ $S_w^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$	$\left(\bar{X}_1-\bar{X}_2\right)\pm t_{1-\alpha/2}(n_1+n_2-2)\cdot S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$
	5、双正态总体均值差，σ_1^2,σ_2^2均未知（大样本）	$Z=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}$ $\sim AN(0,1)$	$(\bar{X}_1-\bar{X}_2)\pm u_{1-\alpha/2}\cdot\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$
	6、、双正态总体均值差。小样本，$\sigma_1^2\neq\sigma_2^2$	$T=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}\dot{\sim}t(f)$	$\left(\bar{X}_1-\bar{X}_2\right)\pm t_{1-\alpha/2}(f)\cdot\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$

		$f=\dfrac{\left(\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(s_1^2/n_1\right)^2}{n_1-1}+\dfrac{\left(s_2^2/n_2\right)^2}{n_2-1}}$	
	7、总体服从Bernoulli分布，大样本	$Z=\dfrac{\bar{X}-p}{\sqrt{\dfrac{p(1-p)}{n}}}\sim AN(0,1)$	$\bar{X}\pm u_{1-\alpha/2}\sqrt{\dfrac{\bar{X}(1-\bar{X})}{n}}$
	8、两个总体独立，分别服从Bernoulli分布，大样本	$Z=\dfrac{\left(\bar{X}_1-\bar{X}_2\right)-\left(p_1-p_2\right)}{\tilde{\sigma}}$ $\sim AN(0,1)$ $\tilde{\sigma}\triangleq\sqrt{\dfrac{\bar{X}_1(1-\bar{X}_1)}{n_1}+\dfrac{\bar{X}_2(1-\bar{X}_2)}{n_2}}$	$\left(\bar{X}_1-\bar{X}_2\right)\pm u_{1-\alpha/2}\cdot\tilde{\sigma}$
总体方差的区间估计	9、总体服从正态分布，μ未知	$\chi^2=\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$	$\left[\dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2}}\ ,\ \dfrac{(n-1)S^2}{\chi^2_{\alpha/2}}\right]$
	10、两个正态总体方差比的区间估计	$F=\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$	$\left[\dfrac{S_1^2/S_2^2}{F_{1-\alpha/2}(n_1-1,n_2-1)},\dfrac{S_1^2/S_2^2}{F_{\alpha/2}(n_1-1,n_2-1)}\right]$

★ **构造枢轴量的一般方法**

设统计量$T\sim F(t,\theta)$（为分布函数），一般可取充分统计量。

由均匀分布的性质知，$F(T,\theta)\sim U[0,1]$，故枢轴量可取为$F(T,\theta)\sim U[0,1]$，

由于

$$P\left(\frac{\alpha}{2}\le F(T,\theta)\le 1-\frac{\alpha}{2}\right)=1-\alpha$$

由此解出 $P(\hat{\theta}_L\le\theta\le\hat{\theta}_U)=1-\alpha$，即得$\theta$**的一个置信度为**$1-\alpha$ **的置信区间**$[\hat{\theta}_L,\hat{\theta}_U]$**.**

例：设样本 $X_1,\cdots,X_n$ 是来自 $N(\mu,\sigma^2),\sigma^2$ 已知，求 μ 的置信度为 $(1-\alpha)\%$ 的置信区间。

解： μ **的充分统计量为** $\bar{X}\sim N(\mu,\frac{\sigma^2}{n})$ **，** $\bar{X}$ **的分布函数为** $F(t,\theta)=\Phi(\frac{t-\mu}{\frac{\sigma}{\sqrt{n}}})$，故

枢轴量可取为 $F(\bar{X},\theta)=\Phi(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}})\sim U[0,1]$

$$P(\frac{\alpha}{2}\le\Phi(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}})\le 1-\frac{\alpha}{2})=1-\alpha$$

即

$$P(\Phi^{-1}(\frac{\alpha}{2})\le\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\le\Phi^{-1}(1-\frac{\alpha}{2}))=1-\alpha$$

$$P(\bar{X}-u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\le\mu\le\bar{X}+u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}})=1-\alpha$$

与以前的结果一致。

注： 对连续型总体，统计量 $T=T(X_1,\cdots,X_n)\sim F(t,\theta)$，若 $F(t,\theta)$ 关于 t,θ 均连续（关于 t 可以不连续），则有

定理：设 $\alpha_1+\alpha_2=\alpha\in(0,1)$，函数 $\theta_1(t),\theta_2(t)$ 满足

$$\begin{aligned}&1-F(t,\theta_1(t))=P(T\ge t)=\alpha_1,\\&F(t+0,\theta_2(t))=P(T\le t)=\alpha_2,\end{aligned}$$

若 $F(t,\theta)$ 关于 θ 严格单减，则 $[\theta_1(T),\theta_2(T)]$ 为**置信度为 $1-\alpha$ 的同等置信区间，**

若 $F(t,\theta)$ 关于 θ 严格单增，则 $[\theta_2(T),\theta_1(T)]$ 为**置信度为 $1-\alpha$ 的同等置信区间。**

补充：

正态总体 (μ,σ^2) 的置信域（二维）

显然，$(\bar{X},S^2)$ 为 (μ,σ^2) 的充分统计量，且 $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\sim N(0,1),\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$ 相

互独立，故可取枢轴量为$(\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}},\frac{(n-1)S^2}{\sigma^2})$，取区域$[c_1,c_2]\times[d_1,d_2]$ 使得

$$P(c_1\le\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\le c_2,d_1\le\frac{(n-1)S^2}{\sigma^2}\le d_2)=1-\alpha$$

即 $P(c_1\le\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\le c_2)P(d_1\le\frac{(n-1)S^2}{\sigma^2}\le d_2)=1-\alpha$

一般可取

$$P(c_1\le\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\le c_2)=\sqrt{1-\alpha}\triangleq 1-\alpha_1$$

$$P(d_1\le\frac{(n-1)S^2}{\sigma^2}\le d_2)=\sqrt{1-\alpha}\triangleq 1-\alpha_1$$

故$c_2=-c_1=u_{1-\frac{\alpha_1}{2}},d_1=\chi^2_{\frac{\alpha_1}{2}}(n-1),d_2=\chi^2_{1-\frac{\alpha_1}{2}}(n-1)$ ，从而(μ,σ^2)的置信域为

$$\{(\mu,\sigma^2):\bar{X}-c_2\frac{\sigma}{\sqrt{n}}\le\mu\le\bar{X}+c_2\frac{\sigma}{\sqrt{n}},\frac{(n-1)S^2}{d_2}\le\sigma^2\le\frac{(n-1)S^2}{d_1}\}$$

例 1　设某糖厂用自动包装机包装糖果，包装好的糖果的重量服从正态分布，且由以往经验知标准差为 0.015（kg）．某日开工后在生产线上抽测 9 袋，得数据 0.497，0.506，0.518，0.524，0.488，0.511，0.510，0.515，0.512（kg），请估计生产线上包装机装箱糖果的期望重量（取$\alpha=0.1$）．

【$\bar{x}\pm u_{1-\alpha/2}\sigma/\sqrt{n}=0.509\pm1.645\times0.015/\sqrt{9}$】

例 2 已知一批零件的长度 X（单位： cm）服从正态分布$N(\mu,1)$， 从中随机地抽取 16 个零件，得到长度的平均值为 40（cm），则μ的置信度为 0.95 的置信区间是________________ .

（注：标准正态分布函数值$\Phi(1.96)=0.975$，$\Phi(1.645)=0.95$）

例 3　假设 0.50, 1.25, 0.80, 2.00 是来自总体 X 的简单样本值. 已知 $Y=\ln X$ 服从正态分布 $N(\mu,1)$.

(1) 求 X 的数学期望 EX (记 EX 为 b);

[$e^{\mu+1/2}$]

(2) 求μ的置信度为 0.95 的置信区间; [-

0.98, 0.98]

(3) 利用上面结果求 b 的置信度为 0.95 的置信区间.

$[e^{-0.48}, e^{1.48}]$

例 4 设 $X_1, X_2, \cdots, X_n$ 是正态总体 $N(\mu,\ 0.9^2)$ 的大小为 n 的简单样本. 为使μ的 0.95 的双侧置信区间长度不超过 1.0, 则样本容量 n 至少应该取多少? 请说明道理.

例 5 为提高某一化学生产过程的得率, 试图采用一种新的催化剂. 为慎重起见, 在实验工厂先进行试验. 设采用原来的催化剂进行了 $n_1 = 8$ 次试验, 得到得率的平均值 $\bar{x}_1 = 91.73$, 样本方差 $s_1{}^2 = 3.89$. 又采用新的催化剂进行了 $n_2 = 8$ 次试验, 得到得率的均值 $\bar{x}_2 = 93.75$, 样本方差 $s_2{}^2 = 4.02$. 假设两总体都可认为服从正态分布, 且方差相等, 试求两总体均值差 $\mu_1 - \mu_2$ 的置信度为 0.95 的单侧置信下限.

第 15 周讲稿 1

2 参数的假设检验

2.1 显著性检验的基本思想与检验的两类错误

★ 两类错误

零假设通常受到保护，而备选假设是当零假设被拒绝后才能被接受。

检验规则：构造一个统计量 $T(X_1,X_2,...,X_3)$，当 H_0 服从某一分布，当 H_0 不成立时，T 的偏大偏小特征。据此，构造拒绝域 W

第一类错误：“弃真”， H_0 为真时，拒绝 H_0； $P\{T \in W \mid H_0 为真\}$

第二类错误：“存伪”， H_0 为不真时，接受 H_0. $P\{T \notin W \mid H_0 为假\}$

势函数： $\beta(\theta) = E_\theta(\delta(X)) = P_\theta\{X \in W\}$ $\delta(X) = \begin{cases} 1, & X \in W. \\ 0, & X \notin W. \end{cases}$

当 $\theta \in \Theta_0$ 时， $\beta(\theta)$ 为犯第一类错误的概率

当 $\theta \in \Theta_1$ 时，$1 - \beta(\theta)$ 为犯第二类错误的概率

★ 假设检验的步骤：

（1） 根据实际问题中所关心的问题，建立原假设 H_0 和备择假设 H_1；

（2） 选择一个合适的统计量 V（要求 H_0 为真时，V 的分布已知），并根据其

取值特点确定一个合适的拒绝域形式；

（3） 由给定的检验水平α，利用关系式

$$P_{H_0}(\{V<V_{\frac{\alpha}{2}}\}\bigcup\{V>V_{1-\frac{\alpha}{2}}\})=\alpha$$

$$P_{H_0}(\{V>V_{1-\alpha}\})=\alpha$$

$$P_{H_0}(\{V<V_{\alpha}\})=\alpha$$

中的某一个，求出水平为α的检验拒绝域.

（4） 根据样本观察值，算出 **V** 的观察值，并根据它作出接受还是拒绝H_0.

一个总体的情况：$X\sim N(\mu,\sigma^2)$

σ^2已知，检验$\mathrm{H}_0{:}\mu=\mu_0\leftrightarrow H_1:\mu\neq\mu_0$：$U=\dfrac{\overline{X}-\mu_0}{\sigma_0\big/\sqrt{n}}\sim N(0,1)$

σ^2未知，检验$\mathrm{H}_0{:}\mu=\mu_0\leftrightarrow H_1:\mu\neq\mu_0$：$T=\dfrac{\overline{X}-\mu_0}{S_n^*\big/\sqrt{n}}\sim t(n-1)$

μ已知，检验$\mathrm{H}_0{:}\sigma^2=\sigma_0^2\leftrightarrow H_1:\sigma^2\neq\sigma_0^2$：$\chi^2=\dfrac{\sum\limits_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}\sim\chi^2(n)$

μ未知，检验$\mathrm{H}_0{:}\sigma^2=\sigma_0^2\leftrightarrow H_1:\sigma^2\neq\sigma_0^2$：$\chi^2=\dfrac{\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}\sim\chi^2(n-1)$

两个总体的情况：$X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$

$\sigma_1^2=\sigma_2^2=\sigma^2$未知时，检验$\mathrm{H}_0{:}\mu_1=\mu_2\leftrightarrow H_1:\mu_1\neq\mu_2$：

$$T=\frac{\overline{X}-\overline{Y}}{\sqrt{(n_1-1)S_{1n_1}^{*2}+(n_2-1)S_{2n_2}^{*2}}}\sqrt{\frac{n_1n_2(n_1+n_2-2)}{n_1+n_2}}\sim t(n_1+n_2-2)$$

μ_1,μ_2未知时，检验$\mathrm{H}_0{:}\sigma_1^2=\sigma_2^2\leftrightarrow H_1:\sigma_1^2\neq\sigma_2^2$：$F=\dfrac{S_{1n_1}^{*2}}{S_{2n_2}^{*2}}\sim F(n_1-1,n_2-1)$

单边检验：举例说明，σ^2已知，检验$\mathrm{H}_0{:}\mu\leq\mu_0\leftrightarrow H_1:\mu>\mu_0$：

构造$U_1=\dfrac{\overline{X}-\mu}{\sigma_0\big/\sqrt{n}}\sim N(0,1)$，给定显著性水平$\alpha$，有$P\{U_1>u_{1-\alpha}\}=\alpha$。当

H_0成立时$U_1=\dfrac{\overline{X}-\mu}{\sigma_0/\sqrt{n}}\ge\dfrac{\overline{X}-\mu_0}{\sigma_0/\sqrt{n}}\overset{def}{\Rightarrow}U$，因此$P\{U>u_{1-\alpha}\}\le P\{U_1>u_{1-\alpha}\}=\alpha$。

故拒绝域为$W=\{U>u_{1-\alpha}\}$

例 7 某厂生产的零件的直径服从正态分布，以往经验知其标准差为 3.6，考虑假设

$$H_0:\mu=68,\quad H_1:\mu\ne 68$$

现按下列方式进行判断：当$|\overline{X}-68|>1$时，拒绝原假设H_0，否则就接受原假设H_0。现在抽取 64 件零件进行检验，则犯第一类错误的概率α

= ________ .【$2(1-\Phi(2.2))$】

例 8　设某糖厂用自动包装机包装糖果，包装好的糖果的重量服从正态分布，且由以往经验知标准差为 0.015（kg）. 某日开工后在生产线上抽测 9 袋，得数据 0.497，0.506，0.518，0.524，0.488，0.511，0.510，0.515，0.512（kg），如果规定包装的每装糖重量为 0.5kg，问由抽测数据能否有 90%的把握判断生产线上包装机工作是否正常？（取 α = 0.1）.

【$H_0=0.5;H_1\ne 0.5,\quad |\dfrac{\overline{x}-0.5}{0.015/\sqrt{9}}|=1.8>z_{\alpha/2}=1.645$，拒绝$H_0$】

2.2 正态总体的参数假设检验

2.2.1 单正态总体参数的双侧检验

1)　H_0：$\mu=\mu_0$　(a) 设σ^2已知：采用 *u* 检验.

(b) 设方差σ^2未知：采用 *t* 检验.

2)　H_0：$\sigma^2=\sigma_0^2$　(a) 设μ已知：采用χ^2检验.

(b) 设μ未知：采用χ^2检验.

2.2.2 两正态总体参数的差异性检验

1) 两个总体均值差$\mu_1-\mu_2$的检验　(a) 设${\sigma_1}^2,{\sigma_2}^2$均为已知，采用 *u* 检验.

(b) 设${\sigma_1}^2={\sigma_2}^2=\sigma^2$，但$\sigma^2$为未知，采用 *t* 检验.

(c) 设${\sigma_1}^2,{\sigma_2}^2$均为未知，且不知它们是否

相等.

2) 方差比的假设检验. H_0: $\sigma_1^2/\sigma_2^2=c$.

2.2.3 一个正态总体参数的单侧检验

〖单侧假设检验，左边假设检验，右边假设检验的对照〗

〖假设检验与区间估计的对照〗

例：设 $X_1,\cdots,X_n$ 是总体 $X\sim N(\mu,\sigma^2)$（σ^2 已知）的一个样本，检验

$$H_0:\mu=\mu_0\leftrightarrow H_1:\mu\neq\mu_0$$

检验统计量为$U=\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\sim N(0,1)$，其拒绝域为

$$W=\{(x_1,x_2,\cdots,x_n):|u|>u_{1-\frac{\alpha}{2}}\}$$

$$=\{(x_1,x_2,\cdots,x_n):|\overline{x}-\mu_0|>u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\}$$

故接受域为

$$W^c=\{(x_1,x_2,\cdots,x_n):|\overline{x}-\mu_0|\le u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\}$$

$$=\{(x_1,x_2,\cdots,x_n):\overline{x}-u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\le\mu_0\le\overline{x}+u_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\}$$

我们发现，将小写改为大写，μ_0改为μ，那么接受域就为参数的置信度为$1-\alpha$的置信区间，即$[\overline{X}-u_{1-\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}},\overline{X}+u_{1-\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}}]$

注：给定样本观测值我们就得到了区间估计的一个实现$[\overline{x}-u_{1-\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}},\overline{x}+u_{1-\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}}]$，而根据区间估计的定义，我们有$1-\alpha$的把握，使得参数真值$\mu$落入这个区间中，但是我们现在要检验的$\mu_0$并没有落入这个区间，那么我们自然会认为$\mu\neq\mu_0$，即这组样本落入了拒绝域。

同样若样本观测值确定，改变原假设中的μ_0，当μ_0取值不同时，我们对原假设的判断也会发生变化。而区间$[\overline{x}-u_{1-\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}},\overline{x}+u_{1-\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}}]$的含义，就

是所有能使原假设被接受的的μ_0集合。

一般有：

结论：设$A(\theta_0)$是$H_0:\theta=\theta_0 \leftrightarrow H_1:\theta \neq \theta_0$的接受域，令

$$C(x_1,x_2,\cdots,x_n)=\{\theta:(x_1,x_2,\cdots,x_n)\in A(\theta_0)\}$$

则$C(X_1,X_2,\cdots,X_n)$为参数θ的置信度为$1-\alpha$的置信区间。

反之，若$C(X_1,X_2,\cdots,X_n)$为参数θ的置信度为$1-\alpha$的置信区间，令

$$A(\theta_0)=\{(x_1,x_2,\cdots,x_n):\theta_0\in C(x_1,x_2,\cdots,x_n)\}$$

则$A(\theta_0)$是$H_0:\theta=\theta_0 \leftrightarrow H_1:\theta \neq \theta_0$的置信水平为$\alpha$的接受域。

- 单参数指数型分布族假设检验问题

对于单参数指数型分布族，若样本$X_1,\cdots,X_n$的联合密度函数为

$$\prod_{i=1}^{n} f(x_i;\theta)=C(\theta)h(x_1,x_2,...,x_n)\exp\{Q(\theta)T(x_1,x_2,...,x_n)\}$$

其中$Q(\theta)$关于θ严格增，我们一般取检验统计量为$T(X_1,X_2,\cdots,X_n)$，考察三类检验问题

$$H_0:\theta\geq\theta_0 \leftrightarrow H_1:\theta<\theta_0;$$
$$H_0:\theta\leq\theta_0 \leftrightarrow H_1:\theta>\theta_0;$$
$$H_0:\theta=\theta_0 \leftrightarrow H_1:\theta\neq\theta_0;$$

在确定拒绝域的形式时，我们需要先求检验统计量的期望$E_\theta T(X_1,X_2,\cdots,X_n)$，如果它是$\theta$的单增函数，那么对于第一个假设而言，当$T(X_1,X_2,\cdots,X_n)$比较小的时候我们会拒绝原假设；类似地可以得到上述三种检验问题的拒绝域形式分别为

$$\{T(x_1,x_2,\cdots,x_n)<c\};$$
$$\{T(x_1,x_2,\cdots,x_n)>c\};$$
$$\{T(x_1,x_2,\cdots,x_n)<c_1\}\cup\{T(x_1,x_2,\cdots,x_n)>c_2\};$$

计算临界值时，我们要使得犯第一类错误的概率小于显著性水平α，即原假设成立的条件下样本落入拒绝域的概率的上确界小于显著性水平。概率的上确界就是$\theta=\theta_0$时的概率，即

$$\sup_{\theta\geq\theta_0}P_\theta\{T(X_1,X_2,\cdots,X_n)<c\}=P_{\theta_0}\{T(X_1,X_2,\cdots,X_n)<c\}$$

（以第一个为例）

若$\theta=\theta_0$时分布完全确定，令上式为α即可求出临界值。

例　设样本$X_1,\cdots,X_n$是来自$E(\frac{1}{\theta})$，检验假设

$$H_0:\theta\geq\theta_0\leftrightarrow H_1:\theta<\theta_0$$

解：样本的分布为

$$f(x_1,x_2,\cdots,x_n;\theta)=\prod_{i=1}^{n}f(x_i;\theta)=(\frac{1}{\theta})^n e^{-\frac{1}{\theta}\sum_{i=1}^{n}x_i}1_{\{x_1,x_2,\cdots,x_n>0\}}$$

取检验统计量为$T(X_1,X_2,\cdots,X_n)=\sum_{i=1}^{n}X_i$。其期望为

$E_\theta T(X_1,X_2,\cdots,X_n)=n\theta$是$\theta$的单增函数，$T=\sum_{i=1}^{n}X_i\sim\Gamma(n,\frac{1}{\theta})$

故拒绝域的形式为$\{T(x_1,x_2,\cdots,x_n)<c\}=\{n\bar{x}<c\}$，犯第一类错误的概率的最大值为$\sup_{\theta\geq\theta_0}P_\theta\{n\bar{X}<c\}=P_{\theta_0}\{n\bar{X}<c\}=\alpha$

由于$T=\sum_{i=1}^{n}X_i\sim\Gamma(n,\frac{1}{\theta})$，故$\frac{2n\bar{X}}{\theta}\sim\Gamma(n,\frac{1}{2})=\chi^2(2n)$，

因此临界值$c=\frac{\theta_0}{2}\chi^2{}_{\alpha}(2n)$。

拒绝域为

$$\{(x_1,x_2,\cdots,x_n):n\overline{x}<\frac{\theta_0}{2}\chi^2{}_{\alpha}(2n)\}$$
$$=\{(x_1,x_2,\cdots,x_n):\frac{2n\overline{x}}{\theta_0}<\chi^2{}_{\alpha}(2n)\}$$

例 9 正常生产条件下，某产品的生产指标 **X～**$N(\mu_0,\sigma_0^2)$，其中 σ_0 **= 0.23. 现在改变了生产工艺，产品的生产指标变为 X′** ～$N(\mu,\sigma^2)$**. 从新工艺产品中任意抽取 10 件，测得均方差为 0.33.** 试在显著性水平α **= 0.05** 下检验 **1)** σ^2无明显变化； **2)** σ^2明显增大.

例 10 某种柴油发动机，每升柴油的运转时间服从正态分布，现测试 **6** 台柴油机，每升柴油的运转时间为 **28，27，31，29，30，27**（分钟），按设计要求每升柴油的运转时间平均应在 **30** 分钟以上，问在显著性水平α **=0.05** 下，这种柴油机是否符合设计要求？

【$H_0:\mu\ge 30;H_1:\mu<30,$ 拒绝域为 $\frac{\overline{x}-\mu_0}{s/\sqrt{n}}<-t_{1-\alpha/2}(n-2)$，

$\overline{x}=28.67,s^2=1.633^2;\frac{\overline{x}-\mu_0}{S/\sqrt{n}}=-2.0>-t_{1-\alpha/2}(n-2)=-2.015$，接受 H_0】

例 11 测得两批电子器件的样品的电阻（欧姆）为

样品 A（*x*）	**0.140**	**0.138**	**0.143**	**0.142**	**0.144**	**0.137**
样品 B (*y*)	**0.135**	**0.140**	**0.142**	**0.136**	**0.138**	**0.140**

设这两批器件的电阻值总体分别服从$N(\mu_1,\sigma_1^2)$ $N(\mu_2,\sigma_2^2)$，且两样本独立.

（1） 检验假设（α **=0.05**）

H_0： $\sigma_1^2=\sigma_2^2$； $H_1:\sigma_1^2\ne\sigma_2^2$

（2） 在（1）的基础上检验（α =0.05）

$$H_0':\ \mu_1=\mu_2;\quad H_1':\mu_1\neq\mu_2$$

【（1） $F=\dfrac{S_1^2}{S_2^2}$，拒绝域为

$$F\geq F_{1-\frac{\alpha}{2}}(n_1-1,n_2-1)\text{或}F\leq F_{\frac{\alpha}{2}}(n_1-1,n_2-1)$$，接受 H_0

（2）拒绝域为，$\dfrac{|\overline{X}-\overline{Y}|}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\geq t_{1-\frac{\alpha}{2}}(n_1+n_2-2)$ ，接受 H_0'

$\bar{x}=0.14067,\bar{y}=0.1385;s_1^2=7.866\times10^{-6},s_1^2=7.1\times10^{-6};s_w=0.002732$

$F_{0.025}(5,5)=7.15,F_{0.975}(5,5)=0.13986;t_{1-\alpha/2}(10)=2.2281$】

例 12 一药厂生产一种新的止痛片，厂方希望验证服用新药片后至开始起作用的时间间隔较原有止痛片至少缩短一半，因此厂方提出需检验假设

$$H_0:\ \mu_1=2\mu_2;\quad H_1:\mu_1>2\mu_2$$

此处，μ_1,μ_2 分别是服用原有止痛片和服用新止痛片后起作用的时间间隔的总体的均值。设两总体均为正态，且方差分别为 σ_1^2，σ_2^2，现分别在两总体中任取一样本 $x_1,x_2,\cdots,x_{n_1}$ 和 $y_1,y_2,\cdots,y_{n_2}$，设两样本独立，试给出上述假设 H_0 的拒绝域（显著性水平为 α）

【检验统计量：$\dfrac{\overline{X}-2\overline{Y}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{4\sigma_2^2}{n_2}}}\sim N(0,1)$，拒绝域：$\dfrac{\overline{X}-2\overline{Y}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{4\sigma_2^2}{n_2}}}\geq u_{1-\alpha}$】

附录 假设检验与区间估计的对照

以单正态总体 $X\sim N(\mu,\sigma^2)$，σ^2 未知为例，考察参数 μ 的区间估计和假设检验问题：

区间估计	假设检验（原假设为 $H_0:\mu=\mu_0$）

枢轴变量：$T=\dfrac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}\sim t(n-1)$	检验统计量： $T=\dfrac{\bar{X}-\mu_0}{\frac{S}{\sqrt{n}}}\sim t(n-1)$
双侧区间：由$\lvert t\rvert<t_{1-\frac{\alpha}{2}}(n-1)$解出$\mu$ 即 $(\bar{X}-t_{1-\frac{\alpha}{2}}(n-1)\dfrac{\sigma}{\sqrt{n}},\ \bar{X}+t_{1-\frac{\alpha}{2}}(n-1)\dfrac{\sigma}{\sqrt{n}})$	双边检验拒绝域： $\lvert t\rvert>t_{1-\frac{\alpha}{2}}(n-1)$ 即$t>t_{1-\frac{\alpha}{2}}(n-1)$或 $t<-t_{1-\frac{\alpha}{2}}(n-1)$
单侧置信上限：$\bar{X}+t_{1-\alpha}(n-1)\dfrac{\sigma}{\sqrt{n}}$ 单侧置信下限：$\bar{X}-t_{1-\alpha}(n-1)\dfrac{\sigma}{\sqrt{n}}$	右边检验$(H_1:\mu>\mu_0)$拒绝域： $t>t_{1-\alpha}(n-1)$； 左边检验$(H_1:\mu<\mu_0)$拒绝域： $t<-t_{1-\alpha}(n-1)$

- ***p* 值**

p和控制犯第一类错的概率α的关系：

Neyman-Pearson 派：α是建立在$H_0\leftrightarrow H_1$体系下的，需要事先设定H_0、H_1，通过事先给定α，即犯第一类错误的概率来划定拒绝域的边界。

Neyman-Pearson 派不会刻意关注p的具体值，他们也不去计算 **p** 值，而是只看观察结果是否落在拒绝域之内。

Fisher 派：p 值，是在原假设正确的前提下，出现观察结果或比之更极端情形的概率，其计算根本不涉及备择假设。**Fisher** 派在计算出 p 值以后，根据其大小判断是否拒绝原假设。认为 p 值的门限可以取 **0.05** 或 **0.01** 等等比较小的数。把这个门限称为显著性水平。

- 定义假设检验的 ***p*** 值是：
-

1) 在原假设为真的前提下出现观察样本以及更极端情况的概率。
2) 利用样本数据能拒绝原假设的最小显著性水平。
3) 观察到的**(**实例样本数据的**)**显著性水平。
4) 表示对原假设的支持程度，是用于确定是否应该拒绝原假设的另一种方法。

- ***p*** 值的计算：

设 T 表示假设检验的检验的统计量，当 **H0** 为真时，可由样本数据计算出该统计量的观测值 C，根据检验统计量 T 的具体分布，可求出 p 值。

（**1**）左侧检验的 p 值为：$p = P(T \le C)$；

（**2**）右侧检验的 p 值为：$p = P(T \ge C)$；

（**3**）双侧检验的 p 值为：检验统计量 T 落在样本统计值 ***C*** 为端点的尾部区域内的概率的 **2** 倍：即

$p = 2P(T \ge C)$ **(**当 **C** 位于分布曲线的右端时**)**；

$p = 2P(T \le C)$ **(**当 **C** 位于分布曲线的左端时**)**

若 ***T*** 服从正态分布和 ***t*** 分布，其分布曲线是关于纵轴对称的，则

$p = P(|T| \ge C)$

- 如何利用 ***p*** 值的检验

  计算出 p 值后，将给定的显著性水平 α 与 p 值比较，就可作出检验的结论：

如果 p 值 $< \alpha$，则在显著性水平 α 下拒绝原假设。

如果 p 值 $\geq\alpha$，则在显著性水平 α 下不拒绝原假设。

★最优势检验

设 $X_1,\cdots,X_n$ 是总体 $X\sim f(x,\theta)(\theta\in\Theta)$ 的一个样本，考虑假设

$$H_0:\theta\in\Theta_0\leftrightarrow H_1:\theta\in\Theta_1$$

$X_1,\cdots,X_n$ 的取值范围 $S_n=W\bigcup W^c$，这里 W 为假设检验的拒绝域。

检验函数：

$$\phi(x_1,\cdots,x_n)=\begin{cases}1, & (x_1,\cdots,x_n)\in W\\ 0, & (x_1,\cdots,x_n)\notin W\end{cases}=1_{\{(x_1,\cdots,x_n)x\in W\}}$$

检验的势函数为

$$g(\theta)=E_\theta(\phi(x_1,\cdots,x_n))=\begin{cases}\alpha(\theta), & \theta\in\Theta_0,\\ 1-\beta(\theta), & \theta\in\Theta_1,\end{cases}$$

定义：当样本容量 n 固定时，检验法 ϕ_1 与 ϕ_2 的势分别为 $g_1(\theta),g_2(\theta)$，水平均为 $\alpha(i.e.\sup\limits_{\theta\in\Theta_0}g_1(\theta)\leq\alpha,\sup\limits_{\theta\in\Theta_0}g_2(\theta)\leq\alpha)$，若 $\forall\theta\in\Theta,$ 均有

$$g_1(\theta)\geq g_2(\theta)$$

则称 ϕ_1 比 ϕ_2 的更有效。

若水平为 α 的检验法中，ϕ_1 比一切水平为 α 的检验法都有效，则称 ϕ_1 是水平为 α 的最优势检验。

一般而言，寻找最优势检验比较困难。有一个比较好的例子是似然比检验。

★似然比检验（简单假设对简单假设）

设 $X_1,\cdots,X_n$ 是总体 $X\sim f(x,\theta)(\theta\in\Theta)$ 的一个样本，考虑简单假设对简单假设的问题

$$H_0:\theta=\theta_0\leftrightarrow H_1:\theta=\theta_1$$

定义似然比

$$\lambda(x_1,x_2,\cdots,x_n)=\frac{L(x_1,x_2,\cdots,x_n,\theta_1)}{L(x_1,x_2,\cdots,x_n,\theta_0)}=\prod_{i=1}^{n}\frac{f(x_i,\theta_1)}{f(x_i,\theta_0)}$$

检验的拒绝域为

$$W=\{(x_1,x_2,\cdots,x_n):\lambda(x_1,x_2,\cdots,x_n)\ge k\}(k\text{ 待定})$$

定理（Neyman-Pearson 引理）设 $X_1,\cdots,X_n$ 是总体 $X\sim f(x,\theta)(\theta\in\Theta)$，对

$$H_0:\theta=\theta_0\leftrightarrow H_1:\theta=\theta_1$$

ϕ 是其似然比检验法，若检验法 ϕ 与 $\tilde{\phi}$ 的势分别为 $g(\theta),\tilde{g}(\theta)$（水平为 α）

满足 $\tilde{g}(\theta_0)\le g(\theta_0)$，则 $\tilde{g}(\theta_1)\le g(\theta_1)$.

即若似然比检验的水平为 α（$i.e.\ g(\theta_0)=\alpha$），则对任一水平为 α 的检验

$\tilde{\phi}$，若 $\tilde{g}(\theta_0)\le\alpha$，则似然比检验 ϕ 为最优势检验。

例 1：设 $X_1,\cdots,X_n$ 是 $N(\mu,\sigma^2)$（σ 已知）的一个样本，检验假设

$$H_0:\mu=\mu_0\leftrightarrow H_1:\mu=\mu_1$$

其中 $\mu_1>\mu_0$

解：似然比

$$\lambda(x_1,x_2,\cdots,x_n)=\frac{L(x_1,x_2,\cdots,x_n,\mu_1)}{L(x_1,x_2,\cdots,x_n,\mu_0)}$$

$$=\frac{(\frac{1}{\sqrt{2\pi}\sigma})^n\exp\{-\sum_{i=1}^{n}\frac{(x_i-\mu_1)^2}{2\sigma^2}\}}{(\frac{1}{\sqrt{2\pi}\sigma})^n\exp\{-\sum_{i=1}^{n}\frac{(x_i-\mu_0)^2}{2\sigma^2}\}}$$

$$=\exp\{n\overline{x}(\frac{\mu_1-\mu_0}{\sigma^2})-\frac{n(\mu_1^2-\mu_0^2)}{2\sigma^2}\}$$

当 H_0 为真时

$$\lambda(x_1,x_2,\cdots,x_n)\ge k\Rightarrow \overline{x}\ge\frac{\ln k+\dfrac{n(\mu_1^2-\mu_0^2)}{2\sigma^2}}{n(\dfrac{\mu_1-\mu_0}{\sigma^2})}$$

$$\Rightarrow\frac{\overline{x}-\mu_0}{\dfrac{\sigma}{\sqrt{n}}}\ge\frac{\sigma\ln k}{\sqrt{n}(\mu_1-\mu_0)}++\frac{\sqrt{n}(\mu_1-\mu_0)}{2\sigma}\triangleq C$$

拒绝域为

$$W=\{(x_1,x_2,\cdots,x_n):\lambda(x_1,x_2,\cdots,x_n)\ge k\}$$

$$=\{(x_1,x_2,\cdots,x_n):\frac{\overline{x}-\mu_0}{\dfrac{\sigma}{\sqrt{n}}}\ge C\}$$

$$=\{(x_1,x_2,\cdots,x_n):\frac{\overline{x}-\mu_0}{\dfrac{\sigma}{\sqrt{n}}}\ge u_{1-\alpha}\}$$

是最优势检验。

例 2：设 $X_1,\cdots,X_n$ 是 $E(\dfrac{1}{\theta})$（σ 已知）的一个样本，检验假设

$$H_0:\theta=\theta_0\leftrightarrow H_1:\theta=\theta_1$$

其中 $\theta_1<\theta_0$

解：似然比

$$\lambda(x_1,x_2,\cdots,x_n)=\frac{L(x_1,x_2,\cdots,x_n,\theta_1)}{L(x_1,x_2,\cdots,x_n,\theta_0)}$$

$$=(\frac{\theta_0}{\theta_1})^n\exp\{-(\frac{1}{\theta_1}-\frac{1}{\theta_0})\sum_{i=1}^{n}x_i\}$$

$$\lambda(x_1,x_2,\cdots,x_n)\ge k$$

$$\Leftrightarrow\frac{2}{\theta_0}\sum_{i=1}^{n}x_i\le\frac{2}{\theta_0}(\frac{1}{\theta_1}-\frac{1}{\theta_0})^{-1}(n\ln\frac{\theta_0}{\theta_1}-\ln k)\triangleq C$$

由于 $T\triangleq\dfrac{2}{\theta_0}\sum_{i=1}^{n}X_i\sim\chi^2(2n)$

最优势检验的拒绝域为$W=\{(x_1,x_2,\cdots,x_n):T\le\chi_\alpha^2(2n)\}$。

★广义似然比检验（复合假设）

$$H_0:\theta\in\Theta_0\leftrightarrow H_1:\theta\in\Theta_1=\Theta-\Theta_0$$

广义似然比

$$\lambda(x_1,x_2,\cdots,x_n)=\frac{\sup\limits_{\theta\in\Theta}L(x_1,x_2,\cdots,x_n,\theta)}{\sup\limits_{\theta\in\Theta_0}(x_1,x_2,\cdots,x_n,\theta)}$$

拒绝域形式：

$$W=\{(x_1,x_2,\cdots,x_n):\lambda(x_1,x_2,\cdots,x_n)\ge C\}$$

$$P(\lambda(X_1,X_2,\cdots,X_n)\ge C)\le\alpha,\forall\theta\in\Theta_0(under\ H_0)$$

例 3 例 1：设$X_1,\cdots,X_n$是$N(\mu,\sigma^2)$（μ,σ均未知）的一个样本，求

$$H_0:\mu=\mu_0\leftrightarrow H_1:\mu=\mu_1$$

的广义似然比检验。

解：记$\theta=(\mu,\sigma),\Theta_0=\{(\mu_0,\sigma):\sigma^2>0\},\Theta=\{(\mu,\sigma):\mu\in R,\sigma^2>0\}$

$$\lambda(x_1,x_2,\cdots,x_n)=\frac{\sup\limits_{\theta\in\Theta}L(x_1,x_2,\cdots,x_n,\theta)}{\sup\limits_{\theta\in\Theta_0}(x_1,x_2,\cdots,x_n,\theta)}$$

$$=\frac{L(x_1,x_2,\cdots,x_n,\overline{x},\frac{n-1}{n}s^2)}{L(x_1,x_2,\cdots,x_n,\mu_0,\sigma^2)}=[\frac{\sum\limits_{i=1}^n(x_i-\mu_0)^2}{\sum\limits_{i=1}^n(x_i-\overline{x})^2}]^{\frac{n}{2}}$$

$$=(1+\frac{t^2}{n-1})^{\frac{n}{2}}\quad t=\frac{\overline{x}-\mu_0}{\frac{s}{\sqrt{n}}}$$

$$\{\lambda(x_1,x_2,\cdots,x_n)\ge C\}=\{|t|\ge d\}$$

$$C=(1+\frac{\alpha^2}{n-1})^{\frac{n}{2}},d=t_{1-\frac{\alpha}{2}}(n-1)$$

$$\Rightarrow W=\{(x_1,x_2,\cdots,x_n):|t|\ge t_{1-\frac{\alpha}{2}}(n-1)\}$$

与以前的双侧检验一致。

★Bayes 检验

设 $X_1,\cdots,X_n$ 是总体 $X \sim f(x,\theta)(\theta\in\Theta)$ 的一个样本，考虑假设

$$H_0:\theta\in\Theta_0 \leftrightarrow H_1:\theta\in\Theta_1=\Theta-\Theta_0$$

Bayes 检验的思想：在得到 θ 的后验分布后，计算此两个假设下的后验概率

$$\alpha_0=P(H_0\,|\,x_1,\cdots,x_n)=P(\theta\in\Theta_0\,|\,x_1,\cdots,x_n)$$
$$\alpha_1=P(H_1\,|\,x_1,\cdots,x_n)=P(\theta\in\Theta_1\,|\,x_1,\cdots,x_n)$$

称 $\dfrac{\alpha_0}{\alpha_1}$ 为后验概率比（Odds）。

检验准则（Jeffreys 准则）

(1) $\dfrac{\alpha_0}{\alpha_1}<1$,则拒绝 H_0，接受 H_1；

(2) $\dfrac{\alpha_0}{\alpha_1}>1$,则接受 H_0；

(1) $\dfrac{\alpha_0}{\alpha_1}<1$,不宜做判断。

先验概率比：

$$\frac{\pi_0}{\pi_1}=\frac{P_\pi(H_0)}{P_\pi(H_1)}=\frac{P_\pi(\theta\in\Theta_0)}{P_\pi(\theta\in\Theta_1)}$$

考虑后验概率比与先验概率比之间的关系，需要引入 H_0 对 H_1 的 Bayes 因子

$B^\pi(x_1,\cdots,x_n)$：

$$B^\pi(x_1,\cdots,x_n)=\frac{\alpha_0/\alpha_1}{\pi_0/\pi_1}$$

实际上，$B^\pi(x_1,\cdots,x_n)=\dfrac{f(x_1,\cdots,x_n\,|\,H_0)}{f(x_1,\cdots,x_n\,|\,H_1)}$

其中 $f(x_1,\cdots,x_n\,|\,H_i)=\int_{\Theta_i} f(x_1,\cdots,x_n\,|\,\theta)\pi_i(\theta)d\theta\ (i=0,1)$ 为边际似然。

这里，我们有

$$\frac{\alpha_0}{\alpha_1}=\frac{P(H_0|x_1,\cdots,x_n)}{P(H_1|x_1,\cdots,x_n)}=\frac{P_\pi(H_0)}{P_\pi(H_1)}B^\pi(x_1,\cdots,x_n)$$

$$=\frac{\pi_0}{\pi_1}B^\pi(x_1,\cdots,x_n)$$

$$\alpha_0=P(H_0|x_1,\cdots,x_n)=[1+\frac{\pi_1}{\pi_0}\frac{1}{B^\pi(x_1,\cdots,x_n)}]^{-1}$$

先验概率确定后，$B^\pi(x_1,\cdots,x_n)$越大，α_0越大，越偏向于接受H_0。

注：Jeffreys 准则（Bayes 因子）

B^π	可信强度（接受H_0的强度）
1:1 至 3:1	不值一提（Barely worth mentioning）
3:1 至 10:1	实质性的（Substantial）
10:1 至 30:1	强（Strong）
30:1 至 100:1	非常强（Very strong）
大于 100:1	决定性的（Decisive）

B^π值小于 1 的，正好反过来（接受H_1的强度）。

情形 1：简单假设对简单假设

$$H_0:\theta=\theta_0\leftrightarrow H_1:\theta=\theta_1$$

此时

$$\alpha_0=P(\theta=\theta_0|x_1,\cdots,x_n)=\frac{\pi_0 f(x_1,\cdots,x_n|\theta_0)}{\pi_0 f(x_1,\cdots,x_n|\theta_0)+\pi_1 f(x_1,\cdots,x_n|\theta_1)}$$

$$\alpha_1=P(\theta=\theta_1|x_1,\cdots,x_n)=\frac{\pi_1 f(x_1,\cdots,x_n|\theta_1)}{\pi_0 f(x_1,\cdots,x_n|\theta_0)+\pi_1 f(x_1,\cdots,x_n|\theta_1)}$$

故

$$\frac{\alpha_0}{\alpha_1}=\frac{\pi_0}{\pi_1}\frac{f(x_1,\cdots,x_n|\theta_0)}{f(x_1,\cdots,x_n|\theta_1)}$$

故 Bayes 因子

$$B^\pi(x_1,\cdots,x_n)=\frac{f(x_1,\cdots,x_n|\theta_0)}{f(x_1,\cdots,x_n|\theta_1)}$$（即似然比）与先验分布无关。

情形 2：复合假设对复合假设

$$H_0:\theta\in\Theta_0\leftrightarrow H_1:\theta\in\Theta_1$$

设θ的先验分布为$\pi(\theta)$，则先验概率

$$\pi_0=P_\pi(H_0)=P_\pi(\theta\in\Theta_0)$$
$$\pi_1=P_\pi(H_1)=P_\pi(\theta\in\Theta_1)$$

先将先验分布$\pi(\theta)$进行分解，使得其支撑集在Θ_0和Θ_1上，记

$$g_0(\theta)=\frac{\pi(\theta)}{\pi_0}1_{\Theta_0}(\theta)$$

$$g_1(\theta)=\frac{\pi(\theta)}{\pi_1}1_{\Theta_1}(\theta)$$

此时先验分布$\pi(\theta)=\pi_0 g_0(\theta)+\pi_1 g_1(\theta)$，于是

$$\frac{\alpha_0}{\alpha_1}=\frac{P(\theta\in\Theta_0\mid x_1,\cdots,x_n)}{P(\theta\in\Theta_1\mid x_1,\cdots,x_n)}=\frac{\int_{\Theta_0}f(x_1,\cdots,x_n\mid\theta)\pi(\theta)d\theta}{\int_{\Theta_1}f(x_1,\cdots,x_n\mid\theta)\pi(\theta)d\theta}$$

$$=\frac{\int_{\Theta_0}f(x_1,\cdots,x_n\mid\theta)\pi_0 g_0(\theta)d\theta}{\int_{\Theta_1}f(x_1,\cdots,x_n\mid\theta)\pi_1 g_1(\theta)d\theta}$$

故有

$$B^\pi(x_1,\cdots,x_n)=\frac{\alpha_0/\alpha_1}{\pi_0/\pi_1}=\frac{\int_{\Theta_0}f(x_1,\cdots,x_n\mid\theta)g_0(\theta)d\theta}{\int_{\Theta_1}f(x_1,\cdots,x_n\mid\theta)g_1(\theta)d\theta}$$（加权似然比）

情形 3：简单假设对复合假设：

$$H_0:\theta=\theta_0\leftrightarrow H_1:\theta\neq\theta_0$$

解：记$\Theta_0=\{\theta_0\},\Theta_1=\Theta-\{\theta_0\}$

$P_\pi(\theta=\theta_0)=\pi_0>0$（依据经验确定）

对于$\Theta_1=\Theta-\{\theta_0\}$上的分布，我们可以给一个正常的密度函数然后进行放缩。这样我们的先验就变成了$\pi(\theta)=\pi_0 1_{\Theta_0}(\theta)+(1-\pi_0)g_1(\theta)1_{\Theta_1}(\theta)$

后验分布为

$$\pi(\theta \mid x_1,\cdots,x_n) \propto \pi_0 f(x_1,\cdots,x_n \mid \theta_0) I_{\Theta_0}(\theta) + \pi_1 g_1(\theta) f(x_1,\cdots,x_n \mid \theta) I_{\Theta_1}(\theta)$$

$$\frac{\alpha_0}{\alpha_1} = \frac{P(\theta = \theta_0 \mid x_1,\cdots,x_n)}{P(\theta \in \Theta_1 \mid x_1,\cdots,x_n)} = \frac{\pi_0 f(x_1,\cdots,x_n \mid \theta)}{\pi_1 \int_{\Theta_1} g_1(\theta) f(x_1,\cdots,x_n \mid \theta) d\theta}$$

$$B^{\pi}(x_1,\cdots,x_n) == \frac{f(x_1,\cdots,x_n \mid \theta)}{\int_{\Theta_1} f(x_1,\cdots,x_n \mid \theta) g_1(\theta) d\theta}$$ 与先验分布无关。

检验可用 Jeffreys 准则（Bayes 因子情形）进行。