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

★ 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$x$的个数$\}$

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

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：** $\tilde{F}_n(x)=\frac{1}{n}\{x_1,\cdots,x_n$ 中小于或等于 $\boldsymbol{x}$ 的个数$\}$

★ **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）

样本均值，样本方差，样本矩

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

（2）

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

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

特别有：

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

例：样本极差的分布密度

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

$$f_R(r)=\int_0^\infty f_{(X_{(1)},X_{(n)})}(u,r+u)du$$

$$=\int_0^\infty n(n-1)[F(r+u)-F(u)]^{n-2}f(u)f(r+u)du,r>0$$

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

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

★ **样本中位数的渐近分布：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 $X_p\dot{\sim}N(x_p,\dfrac{p(1-p)}{n\bullet f^2(x_p)})$

★ **常见统计量：**

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

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

样本矩：$M_k=\dfrac{1}{n}\sum_{i=1}^{n}X_i^k$, $EM_k=EX^k$

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

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

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

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

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** v_α**：** $F(v_\alpha) = P(X \le v_\alpha) = \alpha$

1.5 正态总体样本均值和样本方差的分布

1) 样本均值 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim \boldsymbol{N}(\boldsymbol{\mu}, \sigma^2)$，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） $\chi^2 \equiv \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$，及 $\overline{X}$ 和样本方差 S^2 的独立性.

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

5）$T_{n_1,n_2} \equiv \dfrac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2);\quad S_w^2=\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$

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

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

证明：显然 $n+1$ 维向量 $(\overline{X},X_1-\overline{X},X_2-\overline{X},\cdots,X_n-\overline{X})^T$ 服从 **Gauss** 分布

$$N\left(\begin{pmatrix}\mu^{(1)}\\ \mu^{(2)}\end{pmatrix},\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}\right),$$

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，

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

$$Cov(X_i-\bar{X}, X_j-\bar{X})$$

$$= Cov(X_i, X_j) - Cov(X_i, \bar{X}) - Cov(\bar{X}, X_i) + D\bar{X}$$

$$= \begin{cases} \dfrac{(n-1)\sigma^2}{n}, & i=j \\ -\dfrac{\sigma^2}{n}, & i \neq j \end{cases},$$

从而可知$\overline{X}$和S^2相互独立。

注（**1**）：由于$\sum_{i=1}^{n}\left(\dfrac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(n)$与$(\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}})^2 \sim \chi^2(1)$相互独立，且

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

$$= n(\frac{\overline{X}-\mu}{\sigma})^2 + \frac{(n-1)S^2}{\sigma^2}$$

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

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

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

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， $\overline{X}=\dfrac{1}{n}\sum_{i=1}^{n} X_i$ 和 $S_n{}^2=$

$\dfrac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$

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 $Y_1=\dfrac{1}{6}\sum_{i=1}^{6} X_i$ ， $Y_2=\dfrac{1}{3}\sum_{i=1}^{3} X_{6+i}$， $S^2=$

$\dfrac{1}{2}\sum_{i=7}^{9}\left(X_i-Y_2\right)^2$ 和 $Z=\sqrt{2}\left(Y_1-Y_2\right)/S$. 试证 $Z \sim t(2)$.