数理统计

Didnelpsun

目录

1　统计量

利用期望和方差等数学特征之间的关系进行计算统计量，往往以 $\sum\limits_{i=1}^{n} X_i$ 或类似的形式。

例题：已知总体 X 的期望为 $EX=0$，方差 $DX=\sigma^2$。从总体抽取容量为 n 的简单随机样本，其均值和方差分别为 $\overline{X}$，S^2。记 $S_k^2=\frac{n}{k}\overline{X}^2+\frac{1}{k}S^2$（$k=1,2,3,4$），则 ()。

$A.E(S_1^2)=\sigma^2$　　$B.E(S_2^2)=\sigma^2$

$C.E(S_3^2)=\sigma^2$　　$D.E(S_4^2)=\sigma^2$

解：$E(S_k^2)=E\left(\frac{n}{k}\overline{X}^2+\frac{1}{k}S^2\right)=\frac{n}{k}E\overline{X}^2+\frac{1}{k}E(S^2)=\frac{n}{k}((E\overline{X})^2+D\overline{X})+\frac{1}{k}E(S^2)=\frac{n}{k}\left(0+\frac{\sigma^2}{n}\right)+\frac{1}{k}\sigma^2=\frac{2\sigma^2}{k}$，$\therefore k=2$。

例题：设 X_i 为来自总体 $E(\lambda)$（$\lambda>0$）的简单随机样本，记统计量 $T=\frac{1}{n}\sum\limits_{i=1}^{n}X_i^2$，求 ET。

解：$ET=E\frac{1}{n}\sum\limits_{i=1}^{n}X_i^2=\frac{1}{n}\sum\limits_{i=1}^{n}EX_i^2=\frac{1}{n}\sum\limits_{i=1}^{n}(DX_i+E^2X_i)=\frac{1}{n}\sum\limits_{i=1}^{n}\left(\frac{1}{\lambda^2}+\frac{1}{\lambda^2}\right)=\frac{1}{n}\cdot\frac{2n}{\lambda^2}=\frac{2}{\lambda^2}$。

例题：设 X_i 为来自总体 X 的简单随机样本，而 $X\sim B\left(1,\frac{1}{2}\right)$。记 $\overline{X}=\frac{1}{n}\sum\limits_{i=1}^{n}X_i$，求 $P\left\{\overline{X}=\frac{k}{n}\right\}$。（$0\leqslant k\leqslant n$）

2　三大分布

2.1　χ^2 分布

例题：设 X_1,X_2,X_3,X_4 是来自正态总体 $N(0,4)$ 的简单随机样本，记 $X=a(X_1-2X_2)^2+b(3X_3-4X_4)^2$。求 X 服从 χ^2 分布下的参数与自由度。

解：若 X_1,X_2,X_3,X_4 同一个正态分布，所以 $EX_1=EX_2=EX_3=EX_4=0$，$DX_1=DX_2=DX_3=DX_4=4$。

$E(X_1-2X_2)=EX_1-2EX_2=0$，$D(X_1-2X_2)=DX_1-4DX_2=20$。

$\therefore X_1-2X_2\sim N(0,20)$，同理 $3X_3-4X_4\sim N(0,100)$。

对其标准化：$\frac{X_1-2X_2-0}{\sqrt{20}}\sim N(0,1)$，$\frac{3X_3-4X_4-0}{\sqrt{100}}\sim N(0,1)$。

若要让 X 满足 χ^2 分布，则要将 $a(X_1-2X_2)^2+b(3X_3-4X_4)^2$ 两项标准化。

$\therefore\frac{(X_1-2X_2)^2}{20}+\frac{(3X_3-4X_4)^2}{100}\sim\chi^2(2)$，所以 $a=\frac{1}{20}$，$b=\frac{1}{100}$。

2.2　t 分布

例题：设 $X_1, X_2, \cdots, X_8$ 是来自正态总体 $N(0, 3^2)$ 的简单随机样本，则统计量 $Y = \dfrac{X_1 + X_2 + X_3 + X_4}{\sqrt{X_5^2 + X_6^2 + X_7^2 + X_8^2}}$ 服从什么分布？

解：$\because X_1, \cdots, X_8 \sim N(0, 9)$，$\therefore X_1 + X_2 + X_3 + X_4 \sim N(0, 36)$。

$\therefore \dfrac{X_1 + X_2 + X_3 + X_4 - 0}{6} \sim N(0, 1)$。

$\dfrac{X_5^2 + X_6^2 + X_7^2 + X_8^2}{9} = \left(\dfrac{X_5 - 0}{3}\right)^2 + \left(\dfrac{X_6 - 0}{3}\right)^2 + \left(\dfrac{X_7 - 0}{3}\right)^2 + \left(\dfrac{X_8 - 0}{3}\right)^2 \sim \chi^2(4)$

$\therefore \dfrac{\dfrac{X_1 + X_2 + X_3 + X_4 - 0}{6}}{\sqrt{\dfrac{X_5^2 + X_6^2 + X_7^2 + X_8^2}{9}/4}} = \dfrac{X_1 + X_2 + X_3 + X_4}{\sqrt{X_5^2 + X_6^2 + X_7^2 + X_8^2}} \sim t(4)$。

2.3　F 分布

例题：设 $X_1, X_2, \cdots, X_15$ 是来自正态总体 $N(0, 3^2)$ 的简单随机样本，则统计量 $Y = \dfrac{X_1^2 + X_2^2 + \cdots + X_{10}^2}{2X_{11}^2 + X_{12}^2 + \cdots + X_{15}^2}$ 服从什么分布？

解：$\because \dfrac{X_i - 0}{3} \sim N(0, 1)$，$\left(\dfrac{X_i - 0}{3}\right)^2 = \dfrac{x_i^2}{9} \sim \chi^2(1)$。

$\therefore \dfrac{X_1^2 + X_2^2 + \cdots + X_{10}^2}{9} \sim \chi^2(10)$，$\dfrac{X_{11}^2 + X_{12}^2 + \cdots + X_{15}^2}{9} \sim \chi^2(5)$。

$\therefore \dfrac{\dfrac{X_1^2 + X_2^2 + \cdots + X_{10}^2}{9}/10}{\dfrac{X_{11}^2 + X_{12}^2 + \cdots + X_{15}^2}{9}/5} = \dfrac{X_1^2 + X_2^2 + \cdots + X_{10}^2}{2X_{11}^2 + X_{12}^2 + \cdots + X_{15}^2} = Y \sim F(10, 5)$。

例题：已知 (X, Y) 的概率分布函数为 $f(x, y) = \dfrac{1}{2\pi} e^{-\frac{1}{2}(x^2 + y^2 - 2y + 1)}$，$x, y \in R$，求 $\dfrac{X^2}{(Y-1)^2}$ 的分布。

解：$f(x, y) = \dfrac{1}{2\pi} e^{-\frac{1}{2}(x^2 + y^2 - 2y + 1)} = \dfrac{1}{2\pi} e^{-\frac{1}{2}(x^2 + (y-1)^2)}$，所以根据二维正态分布的形式，得到 $(X, Y) \sim (0, 1; 1, 1; 0)$。

即 $X \sim \Phi(x)$，$Y - 1 \sim \Phi(x)$，$\therefore X^2 \sim \chi^2(1)$，$(Y-1)^2 \sim \chi^2(1)$，$\therefore \dfrac{X^2}{(Y-1)^2} \sim F(1, 1)$。

2.4　函数分布

例题：设随机变量 $X \sim t(n)$，$Y \sim F(1, n)$，常数 C 使得 $P\{X > C\} = 0.6$，求 $P\{Y > C^2\}$。

解：$X \sim t(n)$，则 $X = \dfrac{X_1}{\sqrt{Y_1/n}} \sim t(n)$，其中 $X_1 \sim N(0, 1)$，$Y_1 \sim \chi^2(n)$。

$\therefore X^2 = \frac{X_1^2}{Y_1/n} = \frac{X_1^2/1}{Y_1/n} \sim \frac{\chi^2(1)/1}{\chi^2(n)/n} = F(1,n)$。

又 $P\{Y > C^2\} = 1 - P\{Y \leqslant C^2\}$。$P\{X^2 > C^2\} = 1 - P\{X^2 \leqslant C^2\}$。

又 $P\{X^2 \leqslant C^2\} = P\{-C \leqslant X \leqslant C\}$，根据偶函数性质 $= 0.2$。

$\therefore P\{X^2 > C^2\} = 0.8$。

3 参数估计

3.1 矩估计

4 置信区间

例题：一批零件的长度服从正态分布 $N(\mu,\sigma^2)$，其中 μ,σ^2 均未知。现从中随机抽取 16 个零件，测得样本均值 $\overline{x} = 20cm$，样本标准差为 $s = 1cm$，求 μ 的置信水平为 0.90 的置信区间。

解：σ 未知，所以使用 s 来求置信空间。

置信空间为 $(\overline{X} - t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}, \overline{X} + t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}})$。

已知 $\overline{x} = 20$，$s = 1$，$n = 16$，$\alpha = 1 - 0.90 = 0.1$。

所以置信空间为 $\left(20 - \frac{1}{4}t_{0.05}(15), 20 + \frac{1}{4}t_{0.05}(15)\right)$。

5 假设检验

例题：已知某机器生产出来的零件长度 X（单位：cm）服从正态分布 $N(\mu,\delta^2)$，现从中随意抽取容量为 16 的一个样本，测得样本均值 $\overline{x} = 10$，样本方差 $s^2 = 0.16$，$t_{0.025}(15) = 2.132$。

(1) 求总体均值 μ 置信水平为 0.95 的置信区间。

(2) 在显著性水平 0.05 下检验假设 $H_0 : \mu = 9.7$，$H_1 : \mu \neq 9.7$。

(1) 解：根据公式直接解出置信空间 $(10 - 0.1t_{0.025}(15), 10 + 0.1t_{0.025}(15)) = (9.7868, 10.2132)$。

(2) 解：根据假设 H_0，得到拒绝域 $(-\infty, 9.4868] \cup [9.9132, +\infty)$。

又 $\overline{X} = 10$ 在拒绝域 $[9.9132, +\infty)$ 上，所以假设 H_0 拒绝。

6　两类错误

例题：假定 X 是连续型随机变量，U 是对 X 的一次观测值，关于其概率密度 $f(x)$ 有如下假设：

$$H_0: f(x)=\begin{cases}\dfrac{1}{2}, & 0 \leqslant x \leqslant 2 \\ 0, & \text{其他}\end{cases}\text{，}H_1: f(x)=\begin{cases}\dfrac{x}{2}, & 0 \leqslant x \leqslant 2 \\ 0, & \text{其他}\end{cases}\text{。}$$

检验规则：当事件 $V=\left\{U>\dfrac{3}{2}\right\}$ 出现时，否定假设 H_0，接受 H_1，求犯第一类错误概率和第二类错误概率 $\alpha\beta$。

解：$\alpha=P\left\{U>\dfrac{3}{2}\middle|H_0\right\}=\displaystyle\int_{\frac{3}{2}}^{2}\dfrac{1}{2}\,\mathrm{d}x=\dfrac{1}{4}$。

$\beta=P\left\{U\leqslant\dfrac{3}{2}\middle|H_1\right\}=\displaystyle\int_{0}^{\frac{3}{2}}\dfrac{x}{2}\,\mathrm{d}x=\dfrac{9}{16}$。