数理统计

Didnelpsun

目录

1 统计量 **1**

2 三大分布 **1**

2.1 χ^2 分布 1

2.2 t 分布 2

2.3 F 分布 2

2.4 函数分布 3

3 参数估计 **3**

3.1 矩估计 3

3.1.1 一阶矩 3

3.1.2 二阶矩 3

3.2 最大似然估计 3

4 置信区间 **4**

4.1 方差已知 4

4.2 方差未知 4

5 假设检验 **5**

6 两类错误 **6**

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 $S_k^2=\dfrac{n}{k}\overline{X}^2+\dfrac{1}{k}S^2$（$k=1,2,3,4$），则 ()。

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

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

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

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 $T=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^2$，求 ET。

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

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

解：$\because X\sim B\left(1,\dfrac{1}{2}\right)$，$\therefore\sum\limits_{i=1}^{n}X_i\sim B\left(n,\dfrac{1}{2}\right)$。

$P\left\{\overline{X}=\dfrac{k}{n}\right\}=P\left\{\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i=\dfrac{k}{n}\right\}=P\left\{\sum\limits_{i=1}^{n}X_i=k\right\}=C_n^k\left(\dfrac{1}{2}\right)^k\left(\dfrac{1}{2}\right)^{n-k}$
$=C_n^k\cdot\left(\dfrac{1}{2}\right)^n$。

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

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。

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

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

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

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

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

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

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

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

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

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 $X_1 \sim N(0,1)$，$Y_1 \sim \chi^2(n)$。

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

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

基本方法就是 $EX = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$。

3.1.1 一阶矩

3.1.2 二阶矩

例题：设 X_i 为来自区间 $[-a,a]$ 上均匀分布的总体 X 的简单随机样本，求 a 的矩估计量。

解：首先矩估计就是 $E(X^k) = \frac{1}{n}\sum\limits_{i=1}^{n} X_i^k$。

又对于均匀分布 $X_i \sim U(-a,a)$，$EX = \dfrac{a+b}{2} = 0$，$DX = \dfrac{(b-a)^2}{12} = \dfrac{a^2}{3}$。

所以 EX 不含有 a，使用二阶矩 $EX^2 = DX + E^2X = \dfrac{a^2}{3} = \frac{1}{n}\sum\limits_{i=1}^{n} X_i^2$。

解得 $a = \sqrt{\frac{3}{n}\sum\limits_{i=1}^{n} X_i^2}$。

3.2 最大似然估计

步骤：写出概率函数或密度函数；写出似然函数（代入观测值 x_i 并连乘）；两边取对数；求导数并令为 0。

例题：设随机变量 X 在区间 $[0,\theta]$ 上服从均匀分布，$X_1, X_2, \cdots, X_n$ 是来自 X 的简单随机样本，求 θ 的最大似然估计量 $\hat{\theta}$

解：$X \sim U(0,\theta)$，$f(x)=\begin{cases}\frac{1}{\theta}, & 0<x<\theta \\ 0, & \text{其他}\end{cases}$，$L(\theta)=\begin{cases}\frac{1}{\theta^n}, & 0<x_i<\theta \\ 0, & \text{其他}\end{cases}$。

求 $\hat{\theta}$ 即求 $L(\theta)$ 的最大值，θ 的最小值。又必然 $0<x_i<\theta$。

所以 $\hat{\theta}=\max x_i$，即 θ 的最大似然估计为 $\max\limits_{1\leqslant i\leqslant n} X_i$。

（取最大值而不是最小值是因为为保证所有 x_i 都在定义域上，$0<x_i<\theta$，所以要求 $\theta>\max x_i$）

例题：设 $X_1,X_2,\cdots X_n$ 是来自总体 X 的简单随机样本，X 的概率密度函数 $f(x)=\frac{1}{2\lambda}e^{-\frac{|x|}{\lambda}}$，$x\in R$，$\lambda>0$，求 λ 的最大似然估计量 $\hat{\lambda}$。

解：$\because f(x)=\frac{1}{2\lambda}e^{-\frac{|x|}{\lambda}}$，$\therefore L(\lambda)=\prod\limits_{i=1}^{n}\frac{1}{2\lambda}e^{-\frac{|x|}{\lambda}}=\left(\frac{1}{2\lambda}\right)^n e^{-\frac{1}{\lambda}\sum\limits_{i=1}^{n}|x_i|}$。

$\ln L(\lambda)=-n\ln 2-n\ln\lambda-\frac{1}{\lambda}\sum\limits_{i=1}^{n}|x_i|$，$\frac{\mathrm{d}\ln L(\lambda)}{\mathrm{d}\lambda}=-\frac{n}{\lambda}+\frac{1}{\lambda^2}\sum\limits_{i=1}^{n}|x_i|$。

令 $\frac{\mathrm{d}\ln L(\lambda)}{\mathrm{d}\lambda}=0$，则 $\frac{n}{\lambda}=\frac{1}{\lambda^2}\sum\limits_{i=1}^{n}|x_i|$，解得 $\lambda=\frac{1}{n}\sum\limits_{i=1}^{n}|x_i|$。

即 $\hat{\lambda}=\frac{1}{n}\sum\limits_{i=1}^{n}|X_i|$。

4 置信区间

4.1 方差已知

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 $\left(20-\frac{1}{4}t_{0.05}(15),20+\frac{1}{4}t_{0.05}(15)\right)$。

4.2 方差未知

例题：设某群人的年龄 $X\sim N(\mu,\sigma^2)$，随机了解到五个人的年龄：39，54，61，72，59，求均值 μ 的置信度为 0.95 的置信区间。

解：由于 σ 未知，所以使用样本方差，$\frac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$。

其中置信区间为 $\left(\overline{X}-\frac{S}{\sqrt{n}}t_{0.025}(n-1),\overline{X}+\frac{S}{\sqrt{n}}t_{0.025}(n-1)\right)$。

又 $\overline{x}=\frac{1}{5}(39+54+61+72+59)=57$，$S=\sqrt{\frac{1}{n-1}\sum\limits_{i=1}^{5}(x_i-\overline{x})}=12$。

其中 $t_{0.025}(n-1)=t_{0.025}(4)=2.7764$，所以代入得到 $(42.13,71,87)$。

5 假设检验

例题：设考试成绩服从正态分布，随机抽取 36 位考生成绩，平均分为 66.5 分，标准差为 15 分。在显著性水平 0.05 下是否可以认为这次考试的平均水平为 70 分。

解：首先提出假设 $H_0:\mu=70$，$H_1:\mu\neq 70$。

将 X 使用样本标准差进行标准化：$T=\frac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$。

给定显著性水平 0.05，写出拒绝域 $T<-t_{\frac{\alpha}{2}}(n-1)$ 或 $T>t_{\frac{\alpha}{2}}(n-1)$。

代入计算统计量，$|T|=\left|\frac{\overline{X}-\mu}{S/\sqrt{n}}\right|=\left|\frac{66.5-70}{15/6}\right|=1.4$。

又 $t_{\frac{\alpha}{2}}(n-1)=t_{0.05}(35)=2.0301>1.4$ 不在拒绝域内，所以接受原假设。

即可以认为平均水平为 70 分。

例题：设 $X_1,X_2,\cdots,X_{36}$ 是取自正态总体 $N(\mu,0.04)$ 的简单随机样本，其中 μ 为未知参数，即 $\overline{X}=\frac{1}{36}\sum\limits_{i=1}^{36}X_i$，若对于检验问题 $H_0:\mu\leqslant 0.5$，$H_1:\mu>0.5$ 在显著性水平 $\alpha=0.05$，取得检验拒绝域 $D=\{(x_1,x_2,\cdots,x_{36}):\overline{x}>C\}$，求 C。

解：当 H_0 成立，则 $X\sim N(0.5,0.04)$，$\overline{X}\sim N(0.5,0.04\div 36)=N\left(\frac{1}{2},\frac{1}{900}\right)$。

$\alpha=0.05=P\{$ 拒绝 $H_0|H_0$ 成立 $\}=P\{\overline{X}>C\}=1-P\{\overline{X}\leqslant C\}=1-\Phi((C-0.5)\times 30)=1-\Phi(30C-15)$。

$\therefore\Phi(30C-15)=0.95=\Phi(1.645)$，即 $30C-15=1.645$，$C=0.5548$。

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：

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

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

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