§1 随机样本

总体：研究对象的某项数量指标的值的全体。

个体：总体中的每个元素为个体。

例如：某工厂生产的灯泡的寿命是一个总体，每一个灯泡的寿命是一个个体；某学校男生的身高的全体一个总体，每个男生的身高是一个个体。

定义：设X是具有分布函数F的随机变量，若$X_1,\cdots X_n$是具有同一分布函数F的相互独立的随机变量，则称 $X_1,\cdots X_n$ 为从总体X中得到的容量为n的简单随机样本，简称为样本，其观察值$x_1,\cdots x_n$称为样本值。

返回主目录

由定义知：若 $X_1,\cdots,X_n$ 为X的一个样本，则 $X_1,\cdots,X_n$ 的联合分布函数为：

$$F^*(x_1,\cdots,x_n)=\prod_{i=1}^{n}F(x_i)$$

若设X的概率密度为f，则 $X_1,\cdots,X_n$ 的联合概率密度为：

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

§2 抽样分布

1. 定义：设 $X_1,\cdots X_n$ 为来自总体X的一个样本，$g(X_1,\cdots X_n)$是 $X_1,\cdots X_n$ 的函数，若g是连续函数，且g中不含任何未知参数；

则称$g(X_1,\cdots X_n)$是一个统计量。

设 $x_1,\cdots x_n$是相应于样本$(X_1,\cdots X_n)$的样本值。

则称$g(x_1,\cdots x_n)$是$g(X_1,\cdots X_n)$的观察值。

注：统计量是随机变量。

例1

设 $X_1,\cdots X_n$ 为来自总体 $X\sim N(\mu,\sigma^2)$ 的一个样本，其中μ未知，σ^2已知，问下列随机变量中哪些是统计量

$$\min(X_1,X_2,\cdots,X_n);\ \frac{X_1+X_n}{2};\ \frac{X_1+\cdots+X_n}{n}-\mu\ ;$$

$$\frac{(X_1+X_n)^2}{\sigma^2};\frac{(X_1+\cdots+X_n)-n\mu.}{\sqrt{n}\sigma}.$$

2. 常用的统计量

样本均值：$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$

样本方差：$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2=\frac{1}{n-1}[\sum_{i=1}^{n}X_i^2-n\overline{X}^2]$

样本标准差：$S=\sqrt{S^2}=\sqrt{\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}$

样本k阶(原点)矩：$A_k=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i{}^k\quad k=1,2,\cdots$

样本k阶中心矩：$B_k=\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^k\quad k=1,2,\cdots$

它们的观察值分别为：

$$\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$$

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2=\frac{1}{n-1}[\sum_{i=1}^{n}x_i{}^2-n\overline{x}^2]$$

返回主目录

$$s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2}$$

$$a_k=\frac{1}{n}\sum_{i=1}^{n}x_i^{\ k},\quad k=1,2\cdots$$

$$b_k=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^k,\quad k=1,2\cdots$$

分别称为样本均值、样本方差、样本标准差、样本k阶矩、样本k阶中心矩。

统计量是样本的函数，它是一个随机变量，统计量的分布称为抽样分布。

结论：设 $X_1,\cdots X_n$ 为来自总体 X 的一个样本，

$$EX=\mu,\ DX=\sigma^2,$$

则

P142

$$E\overline{X}=\mu,\ D\overline{X}=\frac{\sigma^2}{n},\ ES^2=\sigma^2$$

$$\mathrm{E}\overline{X}=\frac{1}{n}\sum_{i=1}^{n}EX_i=\frac{1}{n}*n\mu=\mu\quad \mathrm{D}\overline{X}=\frac{1}{n^2}\sum_{i=1}^{n}DX_i=\frac{1}{n^2}*n\sigma^2=\frac{\sigma^2}{n}$$

$$EX_i^2=DX_i+(EX_i)^2=\sigma^2+\mu^2\qquad E\overline{X}^2=D\overline{X}+(E\overline{X})^2=\frac{\sigma^2}{n}+\mu^2$$

$$ES^2=\frac{1}{n-1}[\sum_{i=1}^{n}EX_i^2-nE\overline{X}^2]=\frac{1}{n-1}[n(\sigma^2+\mu^2)-n(\frac{\sigma^2}{n}+\mu^2)]=\sigma^2$$

3. 常用统计量的分布

(1) χ^2 – 分布

设$(X_1,\cdots X_n)$为来自于正态总体$N(0,1)$的样本，则称统计量：

$$\chi^2 = X_1^2 + \cdots + X_n^2$$

所服从的分布为自由度是n的χ^2分布。

记为 $\chi^2 \sim \chi^2(n)$ 概率密度为：$f(y) = \begin{cases} \dfrac{1}{2^{n/2}\Gamma(n/2)} y^{n/2-1} e^{-y/2}, & y > 0 \\ 0, & \text{其它.} \end{cases}$

χ^2分布的性质：

1^0. $\chi_1^2 \sim \chi^2(n_1), \chi_2^2 \sim \chi^2(n_2)$, 且$\chi_1^2$，$\chi_2^2$独立，则有

$$\chi_1^2 + \chi_2^2 \sim \chi^2(n_1 + n_2)$$

返回主目录

2^0. $E\chi^2 = n, \quad D\chi^2 = 2n$

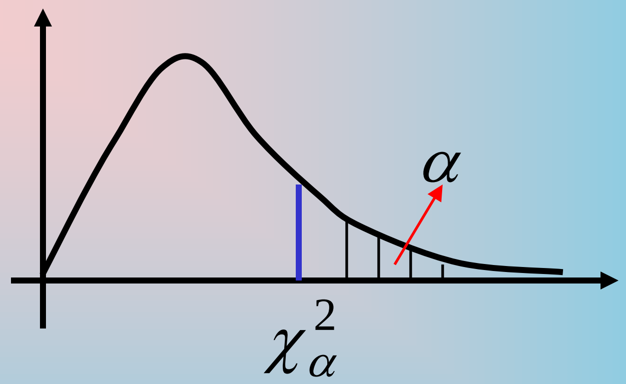
证：$EX_i = 0, \quad DX_i = 1, \quad X_i \sim N(0,1) \quad EX_i^2 = 1,$

$$DX_i^2 = EX_i^4 - (EX_i^2)^2 = 3 - 1 = 2, \quad i = 1,2,\cdots n$$

所以 $E\chi^2 = E(\sum_{i=1}^{n} X_i^2) = \sum_{i=1}^{n} EX_i^2 = n.$

$$D\chi^2 = D(\sum_{i=1}^{n} X_i^2) = \sum_{i=1}^{n} DX_i^2 = 2n.$$

返回主目录



对于给定的$\alpha(0<\alpha<1)$，称满足条件：

$$P\{\chi^2>\chi_\alpha^2(n)\}=\alpha$$

的点$\chi_\alpha^2(n)$为$\chi^2(n)$分布的上α分位点。

当n充分大时，$\chi_\alpha^2(n)\approx\frac{1}{2}(z_\alpha+\sqrt{2n-1})^2$

z_α是标准正态分布的上α分位点。

(2) $t-$分布

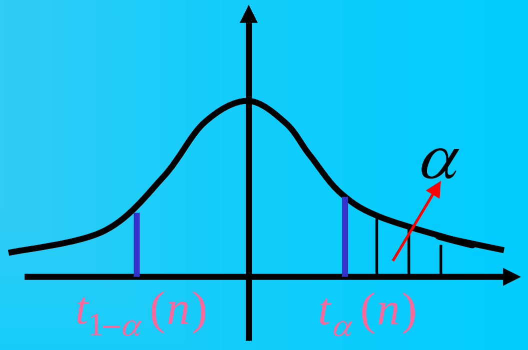
$X\sim N(0,1),\ Y\sim\chi^2(n),\ X,Y$独立，则称随机变量 $T=\dfrac{X}{\sqrt{Y/n}}$

所服从的分布为自由度是n的 $t-$分布，记作$T\sim t(n)$.

对于给定的$\alpha(0<\alpha<1)$，称满足条件：

$$P\{t>t_\alpha(n)\}=\alpha$$

的点$t_\alpha(n)$为t分布的上α分位点。



由概率密度的对称性知：$t_{1-\alpha}(n)=-t_\alpha(n)$

当$n>45$时，$t_\alpha(n)\approx z_\alpha$.

(3) $F-$分布

若 $X \sim \chi^2(n_1), Y \sim \chi^2(n_2)$, X,Y 独立,则称随机变量

$$F=\frac{X/n_1}{Y/n_2}$$

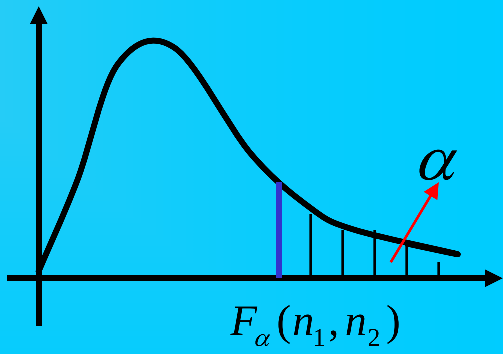
所服从的分布为自由度是n_1,n_2的 $F-$分布，记作$F \sim F(n_1,n_2)$.

若 $F \sim F(n_1,n_2)$,则 $1/F \sim F(n_2,n_1)$.

对于给定的$\alpha(0<\alpha<1)$，称满足条件：

$$P\{F>F_\alpha(n_1,n_2)\}=\alpha$$

的点$F_\alpha(n_1,n_2)$为F分布的上α分位点。



结论： $F_{1-\alpha}(n_1,n_2)=1/F_\alpha(n_2,n_1)$

返回主目录

证明：若$F \sim F(n_1, n_2)$

$$1-\alpha = P\{F > F_{1-\alpha}(n_1, n_2)\} = P\{\frac{1}{F} < \frac{1}{F_{1-\alpha}(n_1, n_2)}\}$$

$$= 1 - P\{\frac{1}{F} \geq \frac{1}{F_{1-\alpha}(n_1, n_2)}\}$$

所以 $P\{\frac{1}{F} > \frac{1}{F_{1-\alpha}(n_1, n_2)}\} = \alpha$ (1)

又由 $1/F \sim F(n_2, n_1)$，所以 $P\{\frac{1}{F} > F_{\alpha}(n_2, n_1)\} = \alpha.$ (2)

比较(1) (2)两式得 $F_{1-\alpha}(n_1, n_2) = \dfrac{1}{F_{\alpha}(n_2, n_1)}$

例： $F_{0.95}(12,9) = \dfrac{1}{F_{0.05}(9,12)} = \dfrac{1}{2.80} = 0.357$

返回主目录

(4) 正态总体的样本均值与样本方差的分布：

定理1. 设$X_1,\cdots,X_n$是总体$N(\mu,\sigma^2)$的样本，$\overline{X},S^2$分别是样本均值与样本方差，则有：

$$(1).\ \overline{X}\sim N(\mu,\frac{\sigma^2}{n}).$$

$$(2).\ \frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$$ 证明见教材$P145$

(3). $\overline{X}$与S^2独立。

定理2. $$\frac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$$

证明： $$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1),\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1).$$

返回主目录

且它们独立。

则由t-分布的定义：$\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\Big/\sqrt{\dfrac{(n-1)S^2}{\sigma^2(n-1)}}\sim t(n-1)$

即：$\dfrac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$

定理3. 设 $X_1,X_2,\cdots,X_{n_1}$ 与 $Y_1,Y_2,\cdots Y_{n_2}$ 分别是具有相同方差的两个正态总体$N(\mu_1,\sigma^2),N(\mu_2,\sigma^2)$的样本，且它们独立。设$\overline{X}=\dfrac{1}{n_1}\sum_{i=1}^{n_1}X_i,\overline{Y}=\dfrac{1}{n_2}\sum_{j=1}^{n_2}Y_j$分别是两个样本的均值。$S_1{}^2=\dfrac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i-\overline{X})^2$

$S_2{}^2=\dfrac{1}{n_2-1}\sum_{j=1}^{n_2}(Y_j-\overline{Y})^2$分别是两个样本的方差；

返回主目录

则有：$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$$

证：$\bar{X}-\bar{Y} \sim N(\mu_1-\mu_2, \frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2})$

所以 $$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{1/n_1+1/n_2}} \sim N(0,1)$$

$\frac{(n_1-1)S_1^2}{\sigma^2} \sim \chi^2(n_1-1), \frac{(n_2-1)S_2^2}{\sigma^2} \sim \chi^2(n_2-1)$，且它们独立。

则 $\frac{(n_1-1)S_1^2}{\sigma^2}+\frac{(n_2-1)S_2^2}{\sigma^2} \sim \chi^2(n_1+n_2-2)$。

返回主目录

由$t-$分布的定义：

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{1/n_1+1/n_2}}\Bigg/\sqrt{\frac{(n_1-1)S_1^2}{\sigma^2}+\frac{(n_2-1)S_2^2}{\sigma^2}/(n_1+n_2-2)}$$

$$\sim t(n_1+n_2-2)$$

即：$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$$

返回主目录

1 给出了总体、个体、样本和统计量的概念，要掌握样本均值和样本方差的计算及基本性质。

2 引进了 χ^2 分布、t分布、F分布的定义，会查表计算。

3 掌握正态总体的某些统计量的分布。

作业：P_{147} 1,3,6,9.