第六章 样本及抽样分布

习 题 课

一、重点与难点

二、主要内容

三、典型例题

一、重点与难点

1.重点

(1) 正态总体某些常用统计量的分布.

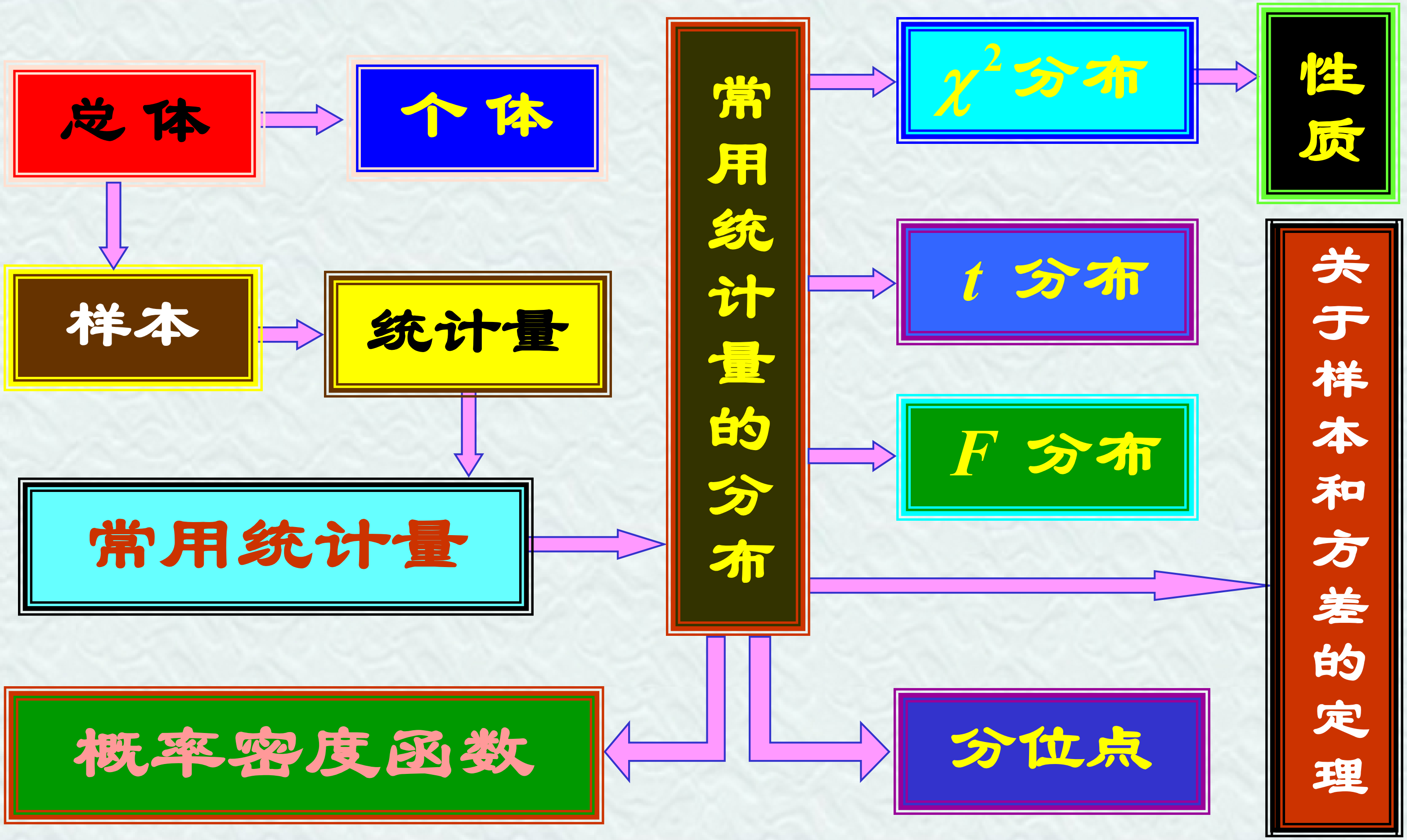
(2) 临界值的查表计算.

2.难点

(1) 几个常用统计量的构造.

(2) 标准正态分布和F分布临界值的查表计算.

二、主要内容



总 体
个 体
样本
统计量
常用统计量
常用统计量的分布
χ^2 分布
性质
t 分布
F 分布
关于样本和方差的定理
概率密度函数
分位点

总体

试验的全部可能的观察值称为总体.

个体

总体中的每个可能观察值称为个体.

样本

设 X 是具有分布函数 F 的随机变量，若 $X_1, X_2, \cdots, X_n$ 是具有同一分布函数 F、相互独立的随机变量，则称 $X_1, X_2, \cdots, X_n$ 为从分布函数 F (或总体 F、或总体 X) 得到的容量为 n 的简单随机样本，简称样本.

统计量

设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的一个样本，$g(X_1, X_2, \cdots, X_n)$ 是 $X_1, X_2, \cdots, X_n$ 的函数，若 g 中不含未知参数，则称 $g(X_1, X_2, \cdots, X_n)$ 是一个统计量.

常用统计量

(1)样本平均值: $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i.$

(2)样本方差:

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

(3)样本标准差:

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}\,.$$

常用统计量

(4)样本 k 阶(原点)矩: $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \ , k = 1, 2, \cdots .$

(5)样本 k 阶中心矩:

$$B_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^k \ , k = 2, 3, \cdots .$$

常用统计量的分布(一)

χ^2分布

设$X_1, X_2, \cdots, X_n$是来自总体$N(0, 1)$的样本，则称统计量 $\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2$ 服从自由度为n的 χ^2 分布，记为 $\chi^2 \sim \chi^2(n)$.

χ^2 分布的性质

性质1 (χ^2分布的可加性)

设 $\chi_1^2 \sim \chi^2(n_1)$, $\chi_2^2 \sim \chi^2(n_2)$, 并且 χ_1^2, χ_2^2 独立, 则 $\chi_1^2 + \chi_2^2 \sim \chi^2(n_1 + n_2)$.

性质2 (χ^2分布的数学期望和方差)

若 $\chi^2 \sim \chi^2(n)$, 则 $E(\chi^2) = n$, $D(\chi^2) = 2n$.

常用统计量的分布(二)

t 分布

设 $X \sim N(0,1)$, $Y \sim \chi^2(n)$, 且 X, Y 独立,

则称随机变量 $t = \dfrac{X}{\sqrt{Y/n}}$ 服从自由度为 n 的 t 分布, 记为 $t \sim t(n)$.

t 分布又称学生氏(Student)分布.

常用统计量的分布(三)

F分布

设$U \sim \chi^2(n_1)$, $V \sim \chi^2(n_2)$, 且U, V 独立,

则称随机变量 $F = \dfrac{U/n_1}{V/n_2}$服从自由度为(n_1, n_2)

的 F 分布, 记为 $F \sim F(n_1, n_2)$.

常用统计量的概率密度函数

$\chi^2(n)$分布的概率密度为

$$f(y)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\Gamma\left(\dfrac{n}{2}\right)}y^{\frac{n}{2}-1}\mathrm{e}^{-\frac{y}{2}}, & y>0,\\ 0 & \text{其他.}\end{cases}$$

常用统计量的概率密度函数

$t(n)$ 分布的概率密度函数为

$$h(t)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}},\quad -\infty<t<+\infty.$$

常用统计量的概率密度函数

$F(n_1, n_2)$分布的概率密度为

$$\psi(y)=\begin{cases}\dfrac{\Gamma\left(\dfrac{n_1+n_2}{2}\right)\left(\dfrac{n_1}{n_2}\right)^{\frac{n_1}{2}} y^{\frac{n_1}{2}-1}}{\Gamma\left(\dfrac{n_1}{2}\right)\Gamma\left(\dfrac{n_2}{2}\right)\left[1+\left(\dfrac{n_1}{n_2}y\right)\right]^{\left(\frac{n_1+n_2}{2}\right)}}, & y>0, \\ 0, & \text{其他}.\end{cases}$$

常用统计量的分布的分位点

χ^2 分布的分位点

对于给定的正数 α, $0<\alpha<1$, 称满足条件

$$P\{\chi^2>\chi^2_\alpha(n)\}=\int_{\chi^2_\alpha(n)}^{+\infty} f(y)\,\mathrm{d}y=\alpha$$

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

常用统计量的分布的分位点

t 分布的分位点

对于给定的 α，$0<\alpha<1$，称满足条件

$$P\{t>t_\alpha(n)\}=\int_{t_\alpha(n)}^{+\infty}h(t)\mathrm{d}t=\alpha$$

的点 $t_\alpha(n)$ 为 $t(n)$ 分布的上 α 分位点.

常用统计量的分布的分位点

F 分布的分位点

对于给定的 α，$0<\alpha<1$，称满足条件

$$P\{F > F_\alpha(n_1, n_2)\} = \int_{F_\alpha(n_1,n_2)}^{+\infty} \psi(y)\mathrm{d}y = \alpha$$

的点 $F_\alpha(n_1, n_2)$ 为 $F(n_1, n_2)$ 分布的上 α 分位点.

F分布的上α分位点具有如下性质：

$$F_{1-\alpha}(n_1, n_2) = \frac{1}{F_\alpha(n_2, n_1)}.$$

关于正态总体的样本和方差的定理

定理一

设 $X_1, X_2, \cdots, X_n$ 是来自正态总体 $N(\mu, \sigma^2)$ 的样本, $\overline{X}$ 是样本均值, 则有 $\overline{X} \sim N(\mu, \sigma^2 / n)$.

定理二

设 $X_1, X_2, \cdots, X_n$ 是总体 $N(\mu, \sigma^2)$ 的样本, $\overline{X}, S^2$ 分别是样本均值和样本方差, 则有

(1) $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$; (2) $\overline{X}$ 与 S^2 独立.

关于正态总体的样本和方差的定理

定理三 设 $X_1, X_2, \cdots, X_n$ 是总体 $N(\mu,\sigma^2)$ 的样本, $\overline{X}, S^2$ 分别是样本均值和样本方差,则有

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

定理四 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正态总体 $N(\mu_1, \sigma^2), N(\mu_2, \sigma^2)$ 的样本，且这两个样本互相独立，设 $\overline{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$，

$\overline{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$ 分别是这两个样本的均值，

$$S_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i - \overline{X})^2,\quad S_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(Y_i - \overline{Y})^2$$

分别是这两个样本的方差，则有

(1) $\dfrac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$;

(2) 当$\sigma_1^2=\sigma_2^2=\sigma^2$时,

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \sim t(n_1+n_2-2),$$

其中 $S_w^2=\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$, $\quad S_w=\sqrt{S_w^2}$.

三、典型例题

例1 设 X 服从 $N(0,1)$, $(X_1, X_2, \cdots, X_6)$为来自总体 X 的简单随机样本,

$$Y = (X_1 + X_2 + X_3)^2 + (X_4 + X_5 + X_6)^2$$

试决定常数 C,使得 CY 服从 χ^2 分布.

解 根据正态分布的性质,

$$X_1 + X_2 + X_3 \sim N(0,3),$$

$$X_4 + X_5 + X_6 \sim N(0,3),$$

则 $\dfrac{X_1+X_2+X_3}{\sqrt{3}} \sim N(0,1),$

$\dfrac{X_4+X_5+X_6}{\sqrt{3}} \sim N(0,1),$

故 $\left(\dfrac{X_1+X_2+X_3}{\sqrt{3}}\right)^2 \sim \chi^2(1),$

$\left(\dfrac{X_4+X_5+X_6}{\sqrt{3}}\right)^2 \sim \chi^2(1),$

因为 $X_1, X_2, \cdots, X_6$ 相互独立及 χ^2 分布的可加性,

$$\left(\frac{X_1+X_2+X_3}{\sqrt{3}}\right)^2+\left(\frac{X_4+X_5+X_6}{\sqrt{3}}\right)^2$$

$$=\frac{1}{3}[(X_1+X_2+X_3)^2+(X_4+X_5+X_6)^2]\sim\chi^2(2),$$

所以 $C=\frac{1}{3}$, CY 服从 χ^2 分布.

例2 设 $\overline{X}_1$ 和 $\overline{X}_2$ 是来自正态总体 $N(\mu,\sigma^2)$ 的容量为 n 的两样本 $(X_{11},X_{12},\cdots,X_{1n})$ 和 $(X_{21},X_{22},\cdots,X_{2n})$ 的样本均值，试确定 n，使得这两个样本均值之差超过 σ 的概率大约为0.01.

解 $\overline{X}_1 \sim N\left(\mu,\dfrac{\sigma^2}{n}\right)$, $\quad \overline{X}_2 \sim N\left(\mu,\dfrac{\sigma^2}{n}\right)$,

则 $\overline{X}_1-\overline{X}_2 \sim N\left(0,\dfrac{2\sigma^2}{n}\right)$,

$$P\{|\overline{X}_1-\overline{X}_2|>\sigma\}=P\left\{\left|\frac{\overline{X}_1-\overline{X}_2}{\sqrt{2/n}\sigma}\right|>\sqrt{\frac{n}{2}}\right\}$$

$$=1-P\left\{\left|\frac{\overline{X}_1-\overline{X}_2}{\sqrt{2/n}\sigma}\right|\le\sqrt{\frac{n}{2}}\right\}$$

$$\approx 1-\left[\Phi\left(\sqrt{\frac{n}{2}}\right)-\Phi\left(-\sqrt{\frac{n}{2}}\right)\right]=2-2\Phi\left(\sqrt{\frac{n}{2}}\right)=0.01,$$

有 $\Phi\left(\sqrt{\frac{n}{2}}\right)\approx 0.995,$ 查标准正态分布表知

$\sqrt{\frac{n}{2}}=2.58,$ 于是 $n=14.$

例3 设总体 $\overline{X} \sim N(\mu,\sigma^2)$,从此总体中取一个容量为 $n=16$ 的样本 $(X_1,X_2,\cdots,X_{16})$,求概率

(1) $P\left\{\dfrac{\sigma^2}{2} \le \dfrac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 \le 2\sigma^2\right\}$;

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

解 (1) 因为 $X_1,X_2,\cdots,X_{16}$ 是来自正态总体的样本,

所以 $\dfrac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2(n)$,

于是 $P\left\{\frac{\sigma^2}{2} \le \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 \le 2\sigma^2\right\}$

$$= P\left\{8 \le \frac{1}{\sigma^2}\sum_{i=1}^{16}(X_i-\mu)^2 \le 32\right\}$$

$$= P\{8 \le \chi^2(16) \le 32\}$$

$$= P\{\chi^2(16) \le 32\} - P\{\chi^2(16) \le 8\}$$

$$= [1 - P\{\chi^2(16) \ge 32\}] - [1 - P\{\chi^2(16) \ge 8\}]$$

$$= 0.94;$$

(2)因为 $\dfrac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\overline{X})^2 \sim \chi^2(n-1)$,

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

$=P\left\{8 \le \dfrac{1}{\sigma^2}\sum_{i=1}^{16}(X_i-\overline{X})^2 \le 32\right\}$

$=P\{8 \le \chi^2(15) \le 32\}$

$=P\{\chi^2(15) \ge 8\}-P\{\chi^2(15) \ge 32\}=0.98$.