第三节　抽样分布

一、分位数的概念

二、χ^2分布

三、t分布

四、F分布

五、抽样分布定理

6.3.1 分位数的概念

1. 分位数定义

定义6.3.1 设连续型随机变量 X 的分布函数为 $F(x)$, 对给定的实数$\alpha(0<\alpha<1)$，如果实数F_α满足

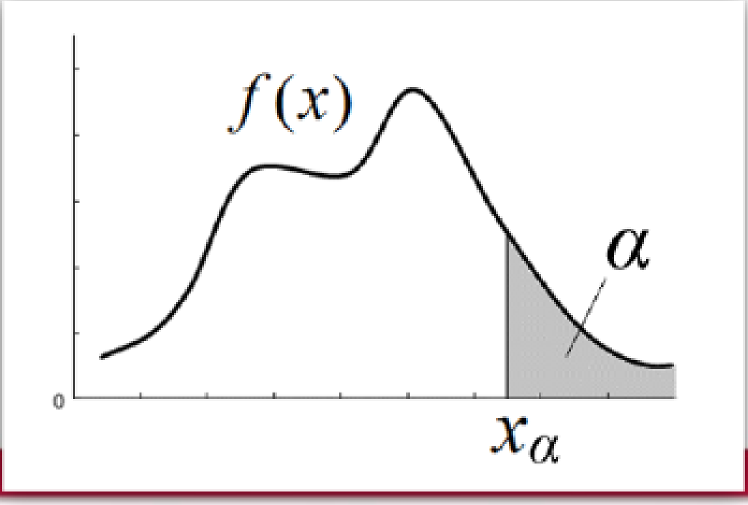
$$P(X>F_\alpha)=\int_{x_\alpha}^{+\infty}f(x)dx=1-F(F_\alpha)=\alpha,$$

则称F_α为随机变量X分布的**水平为α的上侧分位数**或**上α分位数**。

由上述定义可知

$$1-F(F_\alpha)=\alpha$$

$$F(F_\alpha)=1-\alpha$$



6.3.1 分位数的概念

定义6.3.2 设X 是对称分布的连续型随机变量，其分布函数为 $F(x)$, 对给定的实数$\alpha(0<\alpha<1)$，如果正实数$F_{\alpha/2}$满足

$$P(|X|>F_{\alpha/2})=\alpha,$$

则称$F_{\alpha/2}$为随机变量X分布的**水平为α的双侧分位数**或**双侧α分位数**。

$F_{\alpha/2}$即为X分布上的上$\alpha/2$分位数

$$F(F_{\alpha/2})-F(F_{-\alpha/2})=1-\alpha$$

很多统计推断都是基于正态分布的假设，下面讨论三种重要的统计分布：χ^2分布、t分布和F分布。

6.3.2 χ^2分布

2. χ^2定义

定义6.3.3 设随机变量$X_1, X_2, \dots, X_n$独立同$N(0,1)$分布，它们的平方和记为

$$\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2$$

则称χ^2**服从自由度为n的χ^2分布**，记为$\chi^2 \sim \chi^2(n)$.

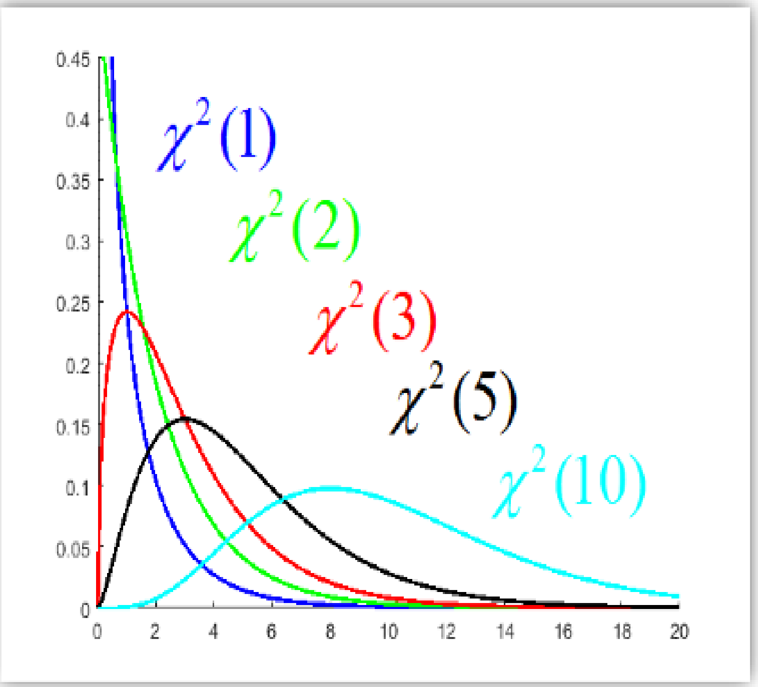
6.3.2 χ^2分布

密度函数为

$$f(x)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}x^{\frac{n-1}{2}}e^{-\frac{1}{2}x}, & x>0\\ 0, & 其他\end{cases}$$

$$\boxed{\Gamma(r)=\int_0^{+\infty}x^{r-1}e^{-x}dx}$$

图像



6.3.2 χ^2分布

性质

6.3.1 数字特征 **设随机变量$X_1, X_2, \dots, X_n$独立同$N(0,1)$分布，$\mathrm{X} = X_1^2 + X_2^2 + \cdots + X_n^2$,则**

$$E(X) = n, D(X) = 2n$$

性质6.3.2 (χ^2分布的可加性或再生性) **设$\mathrm{X} \sim \chi^2(n)$, $\mathrm{Y} \sim \chi^2(m)$,且$X, Y$独立，则$X + Y \sim \chi^2(n+m)$.**

该性质可以推广至有限多个的情形.

6.3.3 t分布

3. t分布定义

定义6.3.4 设随机变量X, Y 相互独立,且

$X \sim N(0,1)$，$Y \sim \chi^2(n)$，令

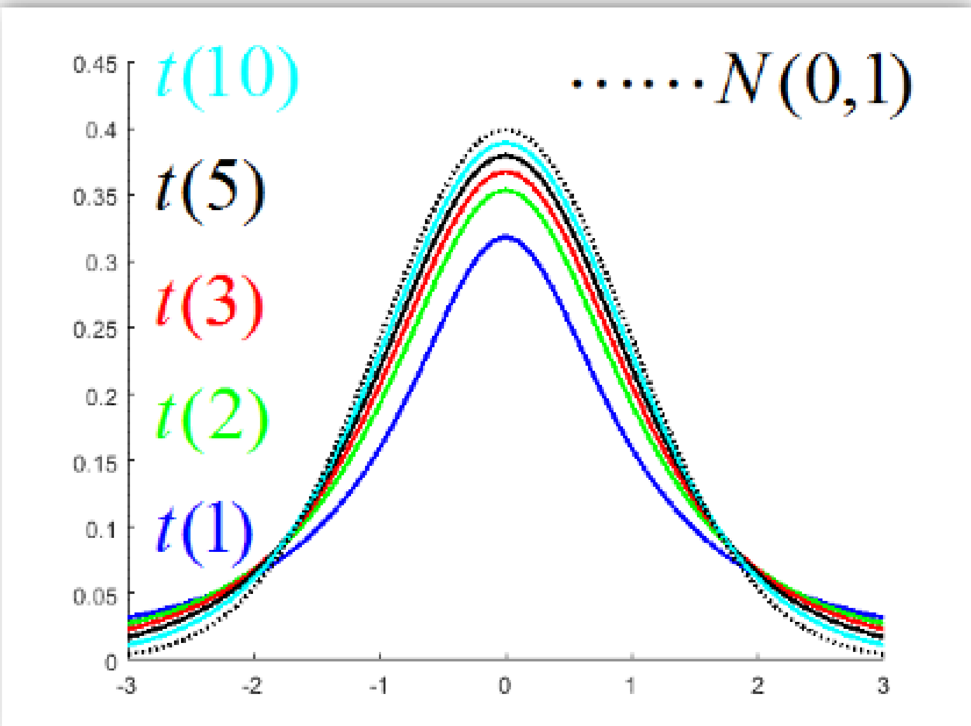
$$t = \frac{X}{\sqrt{Y/n}}$$

则称t 服从自由度为n的t分布，记为$t \sim t(n)$.

6.3.3 t分布

密度函数为

$$f(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}},\ (x\in\mathbf{R}).$$

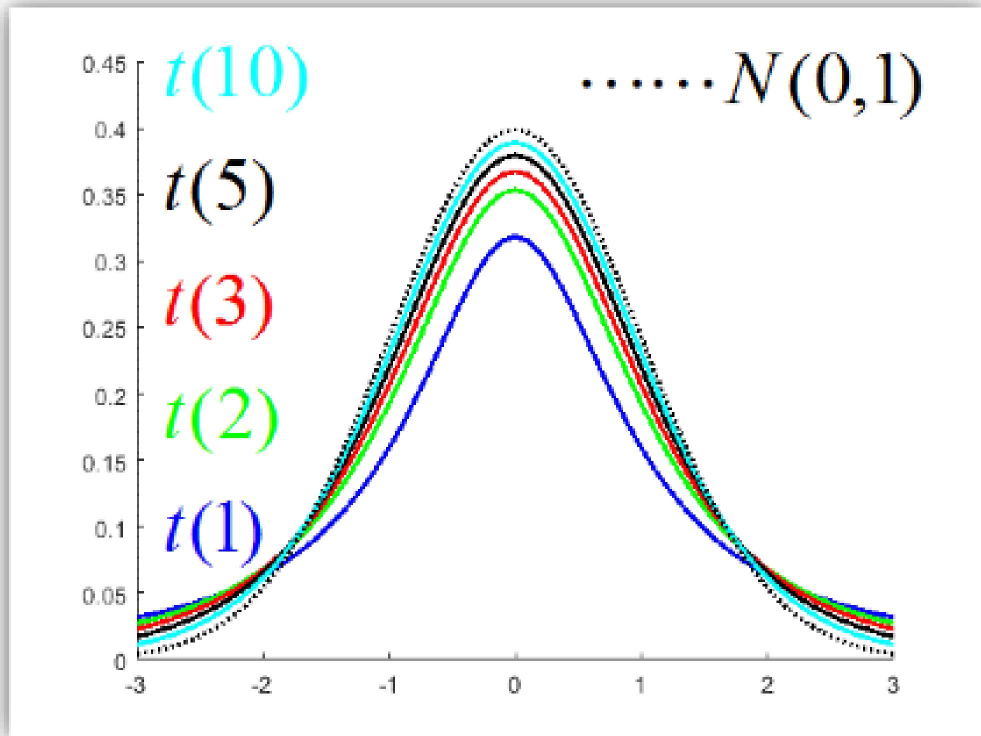


t 分布具有对称的密度曲线

(1) t 分布又称为Student 分布 (学生氏分布).

(2) 可以证明, 当 t 分布的自由度 n 趋向无穷大时, t 分布的极限分布 $N(0, 1)$.

6.3.3 t分布



$$P\big(X > t_\alpha(n)\big) = P\big(X < -t_\alpha(n)\big) = \alpha,$$

$$t_\alpha(n) = -t_{1-\alpha}(n), P\left(|X| > t_{\frac{\alpha}{2}}(n)\right) = \alpha.$$

6.3.4 F分布

4. F分布定义

定义6.3.5 设 $X \sim \chi^2(n_1)$，$Y \sim \chi^2(n_2)$，且二者相互独立，令

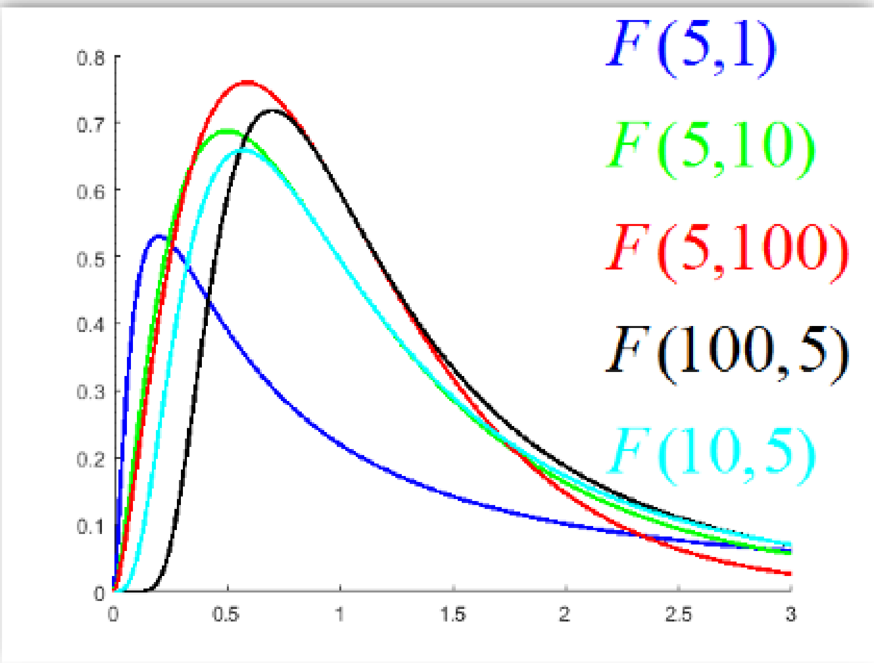
$$F = \frac{X/n_1}{Y/n_2}$$

则称F服从第一自由度为n_1，第二自由度为n_2的F分布，记为 $F \sim F(n_1, n_2)$.

6.3.4 F分布

密度函数为

$$f(x)=\begin{cases}\dfrac{\Gamma\left(\dfrac{n_1+n_2}{2}\right)}{\Gamma\left(\dfrac{n_1}{2}\right)\Gamma\left(\dfrac{n_2}{2}\right)}\left(\dfrac{n_1}{n_2}\right)^{\frac{n_1}{2}}x^{\frac{n_1}{2}-1}\left(1+\dfrac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}}, & x\ge 0,\\ 0, & x<0.\end{cases}$$



6.3.4 F分布

性质

性质6.3.3 若 $F=\dfrac{X/n_1}{Y/n_2}\sim F(n_1,n_2)$, 则$\dfrac{1}{F}=\dfrac{Y/n_2}{X/n_1}\sim F(n_2,n_1)$.

性质6.3.4 若$T\sim t(n)$,则$T^2\sim F(1,n)$.

性质6.3.5

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

5. 抽样分布定理

总体的分布往往是未知的或部分未知的，基于实际问题的需要，有时需对分布类型已知的总体的未知重要数字特征或总体分布中所含的未知参数进行统计推断，这类问题称为参数统计推断。

在参数统计推断问题中，常常需要利用总体的样本构造出合适的统计量，统计学中统称统计量的分布为抽样分布。

6.3.5 抽样分布定理

χ^2分布、 t分布和F分布为讨论正态总体的抽样分布作了必要准备，下面介绍正态总体的样本均值与样本方差的抽样分布以及单个正态总体的抽样分布、两个正态总体的抽样分布定理。

6.3.5 抽样分布定理

回顾

正态分布的可加性:若随机变量$X_1, X_2, \dots, X_n$相互独立，且$X_i \sim N(\mu_i, {\sigma_i}^2), i = 1, 2, \dots n$,则它们的任意线性组合仍服从正态分布，并有

$$\sum_{i=1}^{n} X_i = N(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2)$$

$$c_1X_1 + c_2X_2 + \cdots + c_nX_n \sim N\left(\sum_{i=1}^{n} c_i\,\mu_i, \sum_{i=1}^{n} c_i^2\sigma_i^2\right).$$

6.3.5 抽样分布定理

定理6.3.1 （样本均值和方差的分布 / 费希尔定理） 设总体 $X\sim N(\mu,\sigma^2)$ ，$(X_1,X_2,\ldots,X_n)$是取自总体X的一个容量为n的样本。设

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \qquad S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

分别为样本均值与样本方差，则

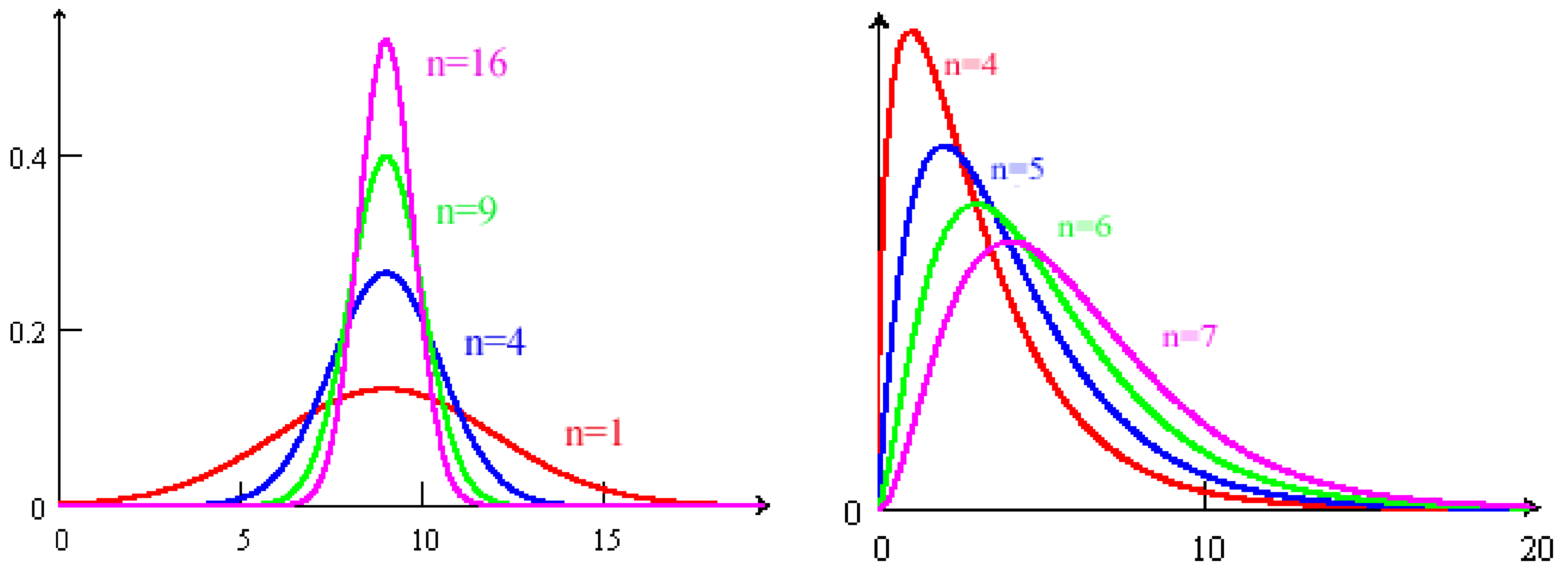
（1）$\bar{X}\sim N(\mu,\frac{\sigma^2}{n})$

(2) $\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$

（3）$\bar{X}$与S^2相互独立

6.3.5 抽样分布定理

（1）$\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$　　（2）$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$　（3）$\overline{X}$与S^2相互独立



n取不同值时样本均值$\overline{X}$和$\frac{(n-1)S^2}{\sigma^2}$的分布

6.3.5 抽样分布定理

定理6.3.2 （单个正态总体的抽样分布） 设$(X_1, X_2, \dots, X_n)$是取自正态总体$X \sim N(\mu, \sigma^2)$ 的样本， $\overline{X}$与S^2分别为该样本均值与样本方差，则

$$(1)\ U = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

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

6.3.5 抽样分布定理

定理6.3.2 （两个正态总体的抽样分布） 设$X \sim N(\mu_1, \sigma_1^2)$ 与Y$\sim N(\mu_2, \sigma_2^2)$是两个相互独立的正态总体。设$(X_1, X_2, \dots, X_n)$是取自总体$X$的样本，$\overline{X}$与$S_1^2$分别为其样本均值与样本方差；$(Y_1, Y_2, \dots, Y_n)$是取自总体$Y$的样本，$\overline{Y}$与$S_2^2$分别为其样本均值与样本方差,则

(1) $U = \dfrac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}} \sim N(0,1)$　　(2) $F = \dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n-1, m-1)$

(3) 当 $\sigma_1^2 = \sigma_2^2 = \sigma^2$时　$T = \dfrac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}}\sqrt{\dfrac{1}{n} + \dfrac{1}{m}}} \sim t(n+m-2)$