统 计 量

设$X_1, X_2, \ldots, X_n$ 是来自总体X的样本，

$g(x_1, x_2, \ldots, x_n)$ 是 n 元连续函数且不含任何未知参数，

则称 $g(X_1, X_2, \ldots, X_n)$ 是统计量. 若 $x_1, x_2, \ldots, x_n$ 是样本观察值，

则称 $g(x_1, x_2, \ldots, x_n)$ 是该统计量的观察值.

常用统计量

设$X_1, X_2, \dots, X_n$ 是来自总体X的样本, $x_1, x_2, \dots, x_n$ 是样本观察值.

样本均值 $\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{\text{观察值}} \bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$　　$\bar{X} - \mu$是统计量吗?

样本方差 $S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2 \xrightarrow{\text{观察值}} s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})^2$

样本标准差 $S = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2} \xrightarrow{\text{观察值}} s = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})^2}$

常用统计量

设$X_1, X_2, \dots, X_n$ 是来自总体X的样本，$x_1, x_2, \dots, x_n$ 是样本观察值.

$X_1^k, X_2^k, \dots, X_n^k$相互独立且与$X^k$ 同分布 $\xRightarrow[\text{数定律}]{\text{辛钦大}}$ $\frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{P} E(X^k)$

样本k阶矩 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \xRightarrow{\text{观察值}} a_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$

样本k阶中心矩 $B_k = \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^k \xRightarrow{\text{观察值}} b_k = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^k$

样本均值 $\bar{X}=\frac{1}{n}\sum_{k=1}^{n}X_k$

设总体X的数学期望和方差存在，且

$$E(X)=\mu \qquad D(X)=\sigma^2,$$

若$X_1,X_2,\ldots,X_n$ 是来自总体X的样本, 则

$$E(\bar{X})=\mu \qquad D(\bar{X})=\frac{\sigma^2}{n}$$

样本方差 $S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X})^2$

设总体X的数学期望和方差存在，且

$$E(X) = \mu, \quad D(X) = \sigma^2,$$

若$X_1, X_2, \ldots, X_n$ 是来自总体X的样本, 则

$$E(S^2) = \sigma^2$$

样本方差 $S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2$

证明： $S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2 = \frac{1}{n-1}\sum_{k=1}^{n}({X_k}^2 - 2X_k\bar{X} + \bar{X}^2)$

$$= \frac{1}{n-1}\left[\sum_{k=1}^{n}{X_k}^2 - \sum_{k=1}^{n}(2X_k\bar{X}) + \sum_{k=1}^{n}\bar{X}^2\right]$$

$$\sum_{k=1}^{n}X_k = n\bar{X}$$

$$= \frac{1}{n-1}\left[\sum_{k=1}^{n}{X_k}^2 - 2\bar{X}\boxed{\sum_{k=1}^{n}X_k} + \sum_{k=1}^{n}\bar{X}^2\right] = \frac{1}{n-1}\left(\sum_{k=1}^{n}{X_k}^2 - n\bar{X}^2\right)$$

$$E(Z^2) = E(Z)^2 + D(Z)$$

$$E(S^2) = \frac{1}{n-1}\left[E\left(\sum_{k=1}^{n} {X_k}^2\right) - nE(\bar{X}^2)\right] = \frac{1}{n-1}\left[\sum_{k=1}^{n} E({X_k}^2) - nE(\bar{X}^2)\right]$$

$$= \frac{1}{n-1}\left[\sum_{k=1}^{n}\{[E(X_k)]^2 + D(X_k)\} - n\{[E(\bar{X})]^2 + D(\bar{X})\}\right]$$

$$= \frac{1}{n-1}\left[\sum_{k=1}^{n}(\mu^2 + \sigma^2) - n\left(\mu^2 + \frac{\sigma^2}{n}\right)\right] = \sigma^2$$

顺序统计量

来自总体X的样本	$X_1, \ldots$	$X_k, \ldots$	X_n
样本的观察值	$x_1, \ldots$	$x_k, \ldots$	x_n
对样本观察值重排	$x_{(1)} \leq \cdots \leq x_{(k)} \leq \cdots \leq x_{(n)}$		
	$X_{(1)} \leq \cdots\ X_{(k)} \leq \cdots \leq\ X_{(n)}$		
样本的顺序统计量	$X_{(1)}, \ldots$	$X_{(k)}, \ldots$	$X_{(n)}$

顺序统计量

设$X_1, X_2, \dots, X_n$ 是来自总体X的样本, 由样本建立 n 个函数:

$$X_{(k)} = X_{(k)}(X_1, X_2, \dots, X_n), k = 1, 2, \dots, n$$

称 $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ 是样本 $X_1, X_2, \dots, X_n$ 的顺序统计量,

$X_{(k)}$是该样本的第k个顺序统计量.

$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$

称 $X_{(1)} = \min(X_1, X_2, \dots, X_n)$ 为最小统计量,

$X_{(n)} = \max(X_1, X_2, \dots, X_n)$ 为最大统计量,

称统计量 $X_{(n)} - X_{(1)}$ 为极差. 对任意实数x,

$$\hat{F}_n(x) = \begin{cases} 0, & x < x_{(1)} \\ \dfrac{1}{n}, & x_{(1)} \le x < x_{(2)} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ \dfrac{k}{n}, & x_{(k)} \le x < x_{(k+1)} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ 1, & x_{(n)} \le x \end{cases}$$

小　结

样本均值 $\bar{X}=\frac{1}{n}\sum_{k=1}^{n}X_k$

样本k阶矩 $A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$

样本方差 $S^2=\frac{1}{n-1}\sum_{k=1}^{n}(X_k-\bar{X})^2$

样本k阶中心矩 $B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k$

样本标准差 $S=\sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(X_k-\bar{X})^2}$

顺序统计量 $X_{(1)},X_{(2)},\dots,X_{(n)}$

$$E(X)=\mu\quad D(X)=\sigma^2\Longrightarrow E(\bar{X})=\mu\quad D(\bar{X})=\frac{\sigma^2}{n}\quad E(S^2)=\sigma^2$$

抽 样 分 布

- 抽样分布
  - χ^2分布
  - t 分布
  - F分布
- 抽样分布的上分位点

抽样分布之 χ^2 分布

设$X_1, X_2, \ldots, X_n$ 相互独立且均服从标准正态分布$N(0, 1)$，称随机变量

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

所服从的分布是自由度为n的χ^2分布，记为$\chi^2 \sim \chi^2(n)$.

特殊的，若 $X \sim N(0, 1)$, 则 $X^2 \sim \chi^2(1)$.

$$\chi^2 \sim \chi^2(n) \longleftrightarrow \chi^2 = \sum_{k=1}^{n} X_k^2 \quad (X_1, X_2, \dots, X_n \ i.i.d. \sim N(0, 1))$$

- 若 $\chi_1^2 \sim \chi^2(m)$，$\chi_2^2 \sim \chi^2(n)$，且χ_1^2与χ_2^2独立，则

$$\chi_1^2 + \chi_2^2 \sim \chi^2(m+n),$$

证明：因为 $\chi_1^2 \sim \chi^2(m)$，故存在m个独立同分布于$N(0, 1)$的随机变量

$X_1, X_2, \dots, X_m$ 使得 $\chi_1^2 = X_1^2 + X_2^2 + \cdots + X_m^2$

同理 $\chi_2^2 \sim \chi^2(n)$，故存在n个独立同分布于$N(0, 1)$的随机变量

$X_{m+1}, X_{m+2}, \dots, X_{m+n}$ 使得 $\chi_2^2 = X_{m+1}^2 + X_{m+2}^2 + \cdots + X_{m+n}^2$

$$\chi^2 \sim \chi^2(n) \Longleftrightarrow \chi^2 = \sum_{k=1}^{n} X_k^2 \quad (X_1, X_2, \dots, X_n \ i.i.d. \sim N(0, 1))$$

χ_1^2与χ_2^2独立 $\Longrightarrow$ 存在$m+n$个独立同分布于$N(0, 1)$的随机变量

$$X_1, X_2, \dots, X_m, X_{m+1}, X_{m+2}, \dots, X_{m+n}$$

使得 $\chi_1^2 + \chi_2^2 = X_1^2 + X_2^2 + \cdots + X_m^2 + X_{m+1}^2 + X_{m+2}^2 + \cdots + X_{m+n}^2$

故 $\chi_1^2 + \chi_2^2 \sim \chi^2(m+n)$

$$\chi^2 \sim \chi^2(n) \Longleftrightarrow \chi^2 = \sum_{k=1}^{n} X_k^2 \quad (X_1, X_2, \ldots, X_n \; i.i.d. \sim N(0,1))$$

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

证明： 因为 $\chi^2 \sim \chi^2(n)$, 故存在n个独立同分布于$N(0, 1)$的随机变量

$X_1, X_2, \ldots, X_n$ 使得 $\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2$

又因为 $E(X_k^2) = D(X_k) + [E(X_k)]^2 = 1, k = 1, 2, \ldots, n$

$$D(X_k^2) = E(X_k^4) - \left[E(X_k^2)\right]^2 = 3 - 1 = 2,\ \ k = 1, 2, \ldots, n$$

所以 $E(\chi^2) = n \quad D(\chi^2) = 2n.$

$$\chi^2 \sim \chi^2(n) \Longleftrightarrow \chi^2 = \sum_{k=1}^{n} X_k^2 \quad (X_1, X_2, \dots, X_n \; i.i.d. \sim N(0,1))$$

例：设总体$X \sim N(0,1)$，$X_1, X_2, \dots, X_6$ 是来自总体X的样本. 又设

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

试确定C, 使得CY服从χ^2分布, 并求此时的$E(CY), D(CY)$.

解：由条件可知 $X_1 + X_2 + X_3 \sim N(0,3)$，$X_4 + X_5 + X_6 \sim N(0,3)$,

则 $$\frac{X_1 + X_2 + X_3}{\sqrt{3}} \sim N(0,1), \quad \frac{X_4 + X_5 + X_6}{\sqrt{3}} \sim N(0,1)$$

$$\chi^2 \sim \chi^2(n) \longleftrightarrow \chi^2 = \sum_{k=1}^{n} X_k^2 \ \ (X_1, X_2, \ldots, X_n \ i.i.d. \sim N(0,1))$$

$\dfrac{X_1+X_2+X_3}{\sqrt{3}}$ 与 $\dfrac{X_4+X_5+X_6}{\sqrt{3}}$ 独立同分布于$N(0,1)$,

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

$$\frac{(X_1+X_2+X_3)^2+(X_4+X_5+X_6)^2}{3} \sim \chi^2(2)$$

即 $\dfrac{Y}{3} \sim \chi^2(2), C=\dfrac{1}{3} \longrightarrow E\left(\dfrac{Y}{3}\right)=2 \quad D\left(\dfrac{Y}{3}\right)=4$

抽样分布之 t 分布

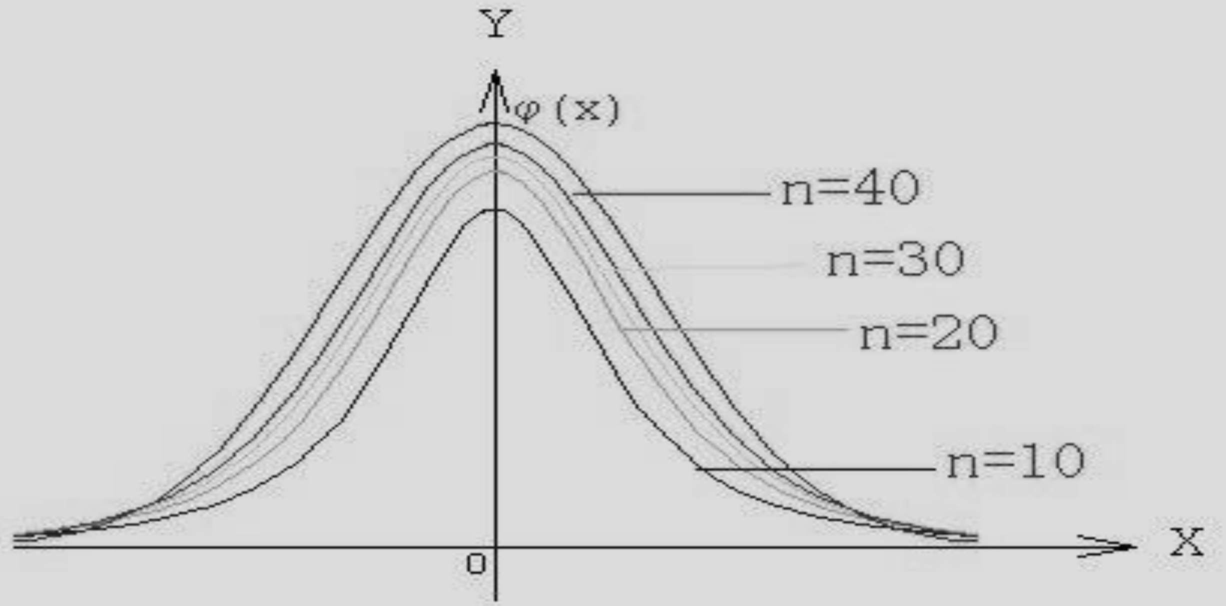
设随机变量 $X \sim N\,(0,1)$, $Y \sim \chi^2(n)$, 并且X与Y独立，则称随机变量

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

所服从的分布是自由度为n的t分布，记为$T \sim t(n)$.

t 分布的密度函数$h(t)$的图形

$$\lim_{n\to\infty} h(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$



抽样分布之 F 分布

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

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

所服从的分布是自由度为n_1, n_2的F分布，记为$F \sim F(n_1, n_2)$.

$$F\sim F(n_1,n_2) \Longleftrightarrow F=\frac{X/n_1}{Y/n_2} \quad (X\sim\chi^2(n_1)\, Y\sim\chi^2(n_2)\ X与Y独立)$$

- 若 $F\sim F(n_1,n_2)$, 则 $\dfrac{1}{F}\sim F(n_2,n_1)$.

证明：因为 $F\sim F(n_1,n_2)$, 故存在相互独立的两个随机变量X, Y,

其中 $X\sim\chi^2(n_1)$, $Y\sim\chi^2(n_2)$, 使得 $F=\dfrac{X/n_1}{Y/n_2}$

故 $\dfrac{1}{F}=\dfrac{Y/n_2}{X/n_1}$ 即 $\dfrac{1}{F}\sim F(n_2,n_1)$.

$$F \sim F(n_1, n_2) \Longleftrightarrow F = \frac{X/n_1}{Y/n_2} \quad (X \sim \chi^2(n_1)\ Y \sim \chi^2(n_2)\ X\text{与}Y\text{独立})$$

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

$$\left.\begin{array}{r} X \sim N(0,1) \\ Y \sim \chi^2(n) \\ X\text{与}Y\text{相互独立} \end{array}\right\} \Longrightarrow T = \frac{X}{\sqrt{Y/n}} \Longrightarrow T^2 = \frac{X^2}{Y/n}$$

$$\left.\begin{array}{r} T^2 = \dfrac{X^2}{Y/n} \\ X^2 \sim \chi^2(1) \\ X^2\text{与}Y\text{相互独立} \end{array}\right\} \Longrightarrow T^2 \sim F(1, n).$$

$$T \sim t(n) \Longleftrightarrow T = \frac{X}{\sqrt{Y/n}} \quad (X \sim N(0,1)\ \ Y \sim \chi^2(n)\ \ X\text{与}Y\text{独立})$$

上分位点

设X是连续型随机变量，α是给定常数$(0 < \alpha < 1)$，若实数u_α满足

$$P(X > u_\alpha) = \alpha,$$

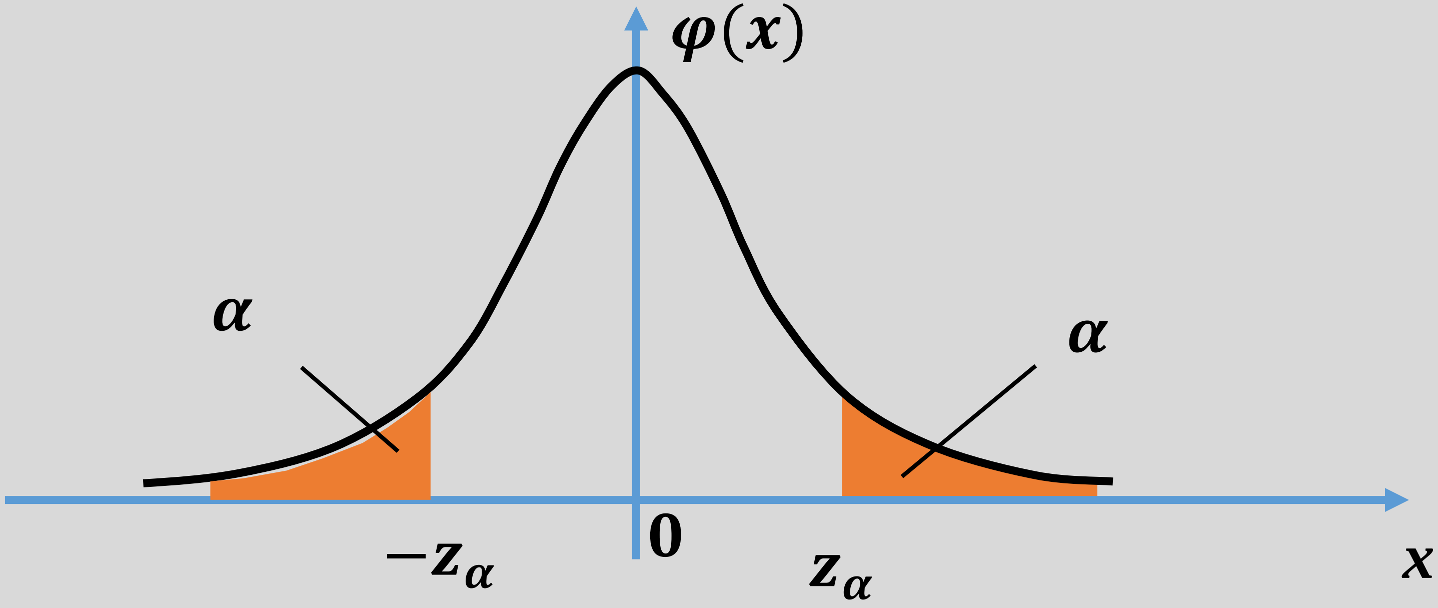
则称u_α是X或其分布的上α分位数.

设 $X \sim N(0,1)$, 对给定的常数$\alpha(0 < \alpha < 1)$，若实数z_α满足

$$P(X > z_\alpha) = \alpha,$$

则称z_α是X或标准正态分布$N(0,1)$的上α分位数.

$$P(X \geq z_\alpha) = \alpha$$



$$P(X > z_\alpha) = \alpha \implies P(X \geq -z_\alpha) = 1 - \alpha$$

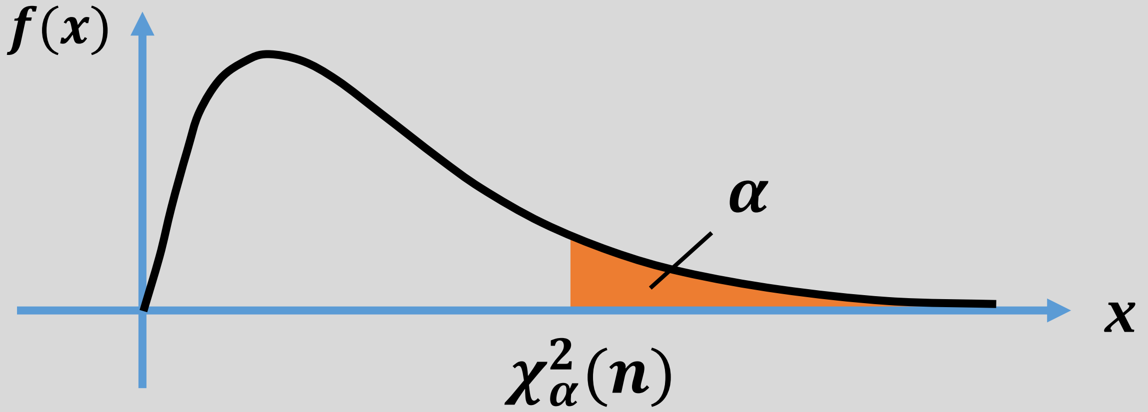
$$\implies z_{1-\alpha} = -z_\alpha$$

χ^2分布的上α分位点

设 $X \sim \chi^2(n)$, 对给定的常数$\alpha(0 < \alpha < 1)$, 若实数$\chi^2_\alpha(n)$满足

$$\boxed{P\left(X > \chi^2_\alpha(n)\right) = \alpha,}$$

则称$\chi^2_\alpha(n)$是$\chi^2(n)$分布的上α分位数.

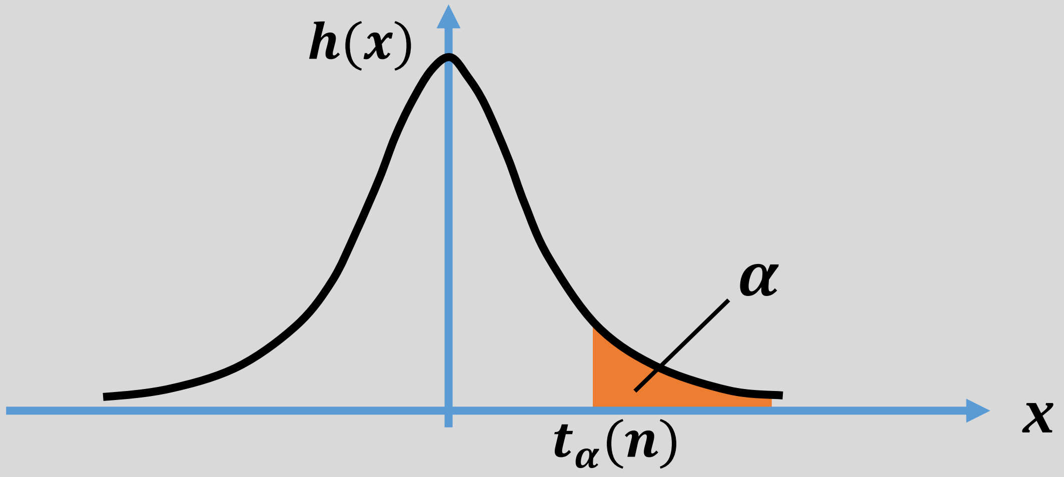


t分布的上α分位点

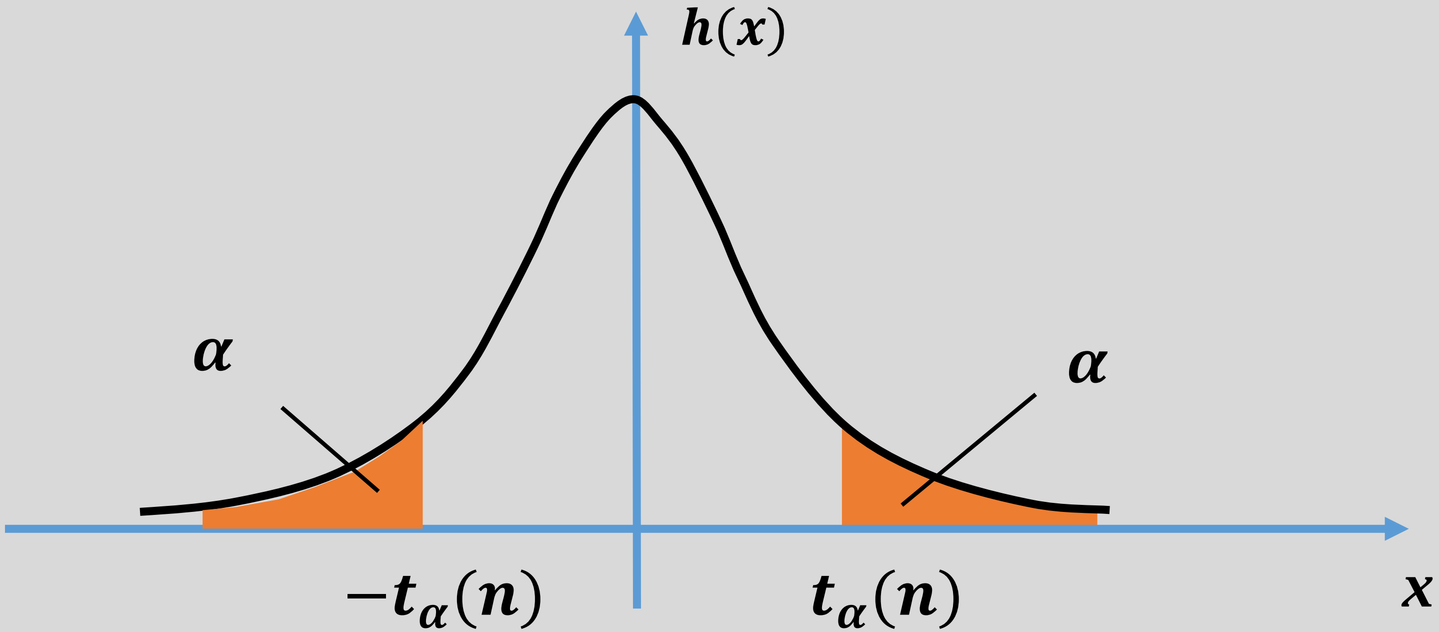
设 $X \sim t(n)$, 对给定的常数$\alpha(0 < \alpha < 1)$，若实数$t_\alpha(n)$满足

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

则称$t_\alpha(n)$是$t(n)$分布的上α分位数.



$$P(X > t_\alpha(n)) = \alpha, \qquad (0 < \alpha < 1)$$



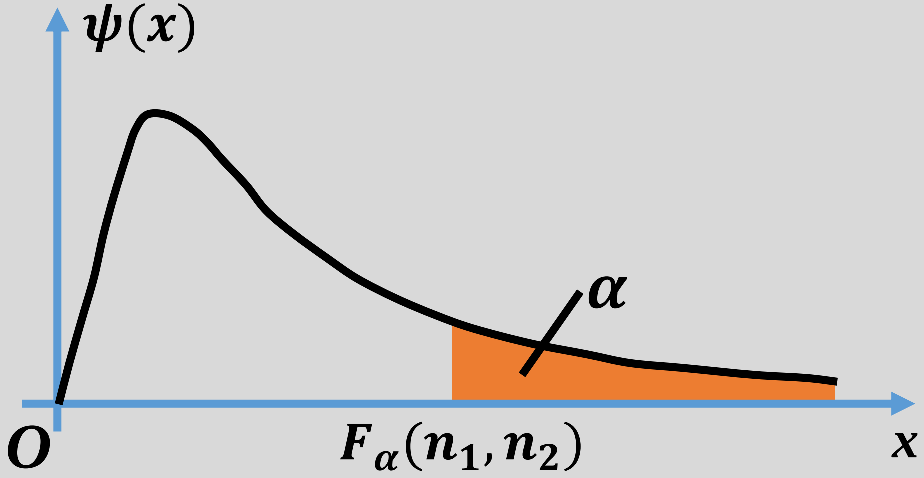
$$P(X > -t_\alpha(n)) = 1 - \alpha \Longrightarrow t_{1-\alpha}(n) = -t_\alpha(n)$$

F分布的上α分位点

设 $X\sim F(n_1,n_2)$, 对给定的常数$\alpha(0<\alpha<1)$, 若实数$F_\alpha(n_1,n_2)$满足

$$P(X>F_\alpha(n_1,n_2))=\alpha,$$

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



$$P(X > F_\alpha(n_1,n_2)) = \alpha, \qquad (0 < \alpha < 1)$$

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

证明：设 $X \sim F(n_1,n_2)$, 则 $\frac{1}{X} \sim F(n_2,n_1)$.

$$\Longrightarrow P\left(\frac{1}{X} > F_{1-\alpha}(n_2,n_1)\right) = 1-\alpha$$

$$\Longrightarrow P\left(X \le \frac{1}{F_{1-\alpha}(n_2,n_1)}\right) = 1-\alpha$$

$$\Longrightarrow P\left(X > \frac{1}{F_{1-\alpha}(n_2,n_1)}\right) = \alpha \Longrightarrow F_\alpha(n_1,n_2) = \frac{1}{F_{1-\alpha}(n_2,n_1)}$$

小　结

- 抽样分布
  - χ^2分布
  - t 分布
  - F分布

- 抽样分布的上分位点