【例 1.7(P_{13})】从某地区随机抽取 50 户农民, 调查其年收入情况, 得到下列数据 (每户人均元):

924	800	916	704	870	1040	824	690	574	490
972	988	1266	684	764	940	408	804	610	852
602	754	788	962	704	712	854	888	768	848
882	1192	820	878	614	846	746	828	792	872
696	644	926	808	1010	728	742	850	864	738

试对该地区农民的收入水平和贫富悬殊程度做个大致分析.

一、统计量

1.定义: 若$T(X_1, X_2, \cdots, X_n)$是样本$X_1, X_2, \cdots, X_n$的不含未知参数的函数, 则称$T(X_1, X_2, \cdots, X_n)$是一统计量.

$$\begin{array}{ccc}
\text{样本}X_1, X_2, \ldots, X_n & \xrightarrow{\text{加 工}} & \boxed{\text{统计量}T(X_1, X_2, \ldots, X_n)} \\
\downarrow & & \downarrow \\
\text{样本观察值}x_1, x_2, \ldots, x_n & & \text{统计量的观察值}T(x_1, x_2, \ldots, x_n)
\end{array}$$

【例1.10(P14)】设$X_1, X_2, \cdots, X_n$是来自总体$X \sim N(\mu, \sigma^2)$的样本，其中μ与σ为未知参数，判断下列函数哪些是统计量:

$$X_1, \qquad \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2, \qquad \frac{X_1}{\sigma}$$

【例1.11(P14)】样本观察值$x_1, x_2, \cdots, x_n$落在某指定区间Δ中的频数n_Δ、频率$f_\Delta = \dfrac{n_\Delta}{n}$，以及经验分布函数$F_n(x)$($x$固定)是否是统计量?

2.几个常用统计量(例1.8, 例1.9(P14)):

(1) 设$X_1,X_2,\cdots,X_n$为总体X的样本, 则

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

样本方差 $$S^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2 \quad 或 \quad S_*^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

样本标准差 $$S=\sqrt{S^2} \quad 或 \quad S_*=\sqrt{S_*^2}$$

样本l阶原点矩 $$A_l=\frac{1}{n}\sum_{i=1}^{n}X_i^l,\quad l=1,2,\cdots$$

(2) 设$(X_1,Y_1),(X_2,Y_2),\cdots,(X_n,Y_n)$为二维总体$(X,Y)$的样本, 则

样本协方差 $$S_{XY}=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})$$

样本相关系数 $$\hat{\rho}_{XY}=\frac{S_{XY}}{\sqrt{S_{XX}}\sqrt{S_{YY}}}$$

为更好地理解下列统计量的形式与含义，设二维总体(X,Y)为离散型，其联合分布律与X的边缘分布律如下:

$X \backslash Y$	y_1	$\cdots$	y_k	$\cdots$	$P(X=x_j)\triangleq p_j$
x_1	p_{11}	$\cdots$	p_{1k}	$\cdots$	p_1
$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$
x_j	p_{j1}	$\cdots$	p_{jk}	$\cdots$	p_j
$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$

$(X_1,Y_1),(X_2,Y_2),\cdots,(X_n,Y_n)$为$(X,Y)$的样本. (注意$P(X=x_j)\triangleq p_j$只在此处不引起混淆的情况下才如此记)

表1 总体与样本下的五个概念对比

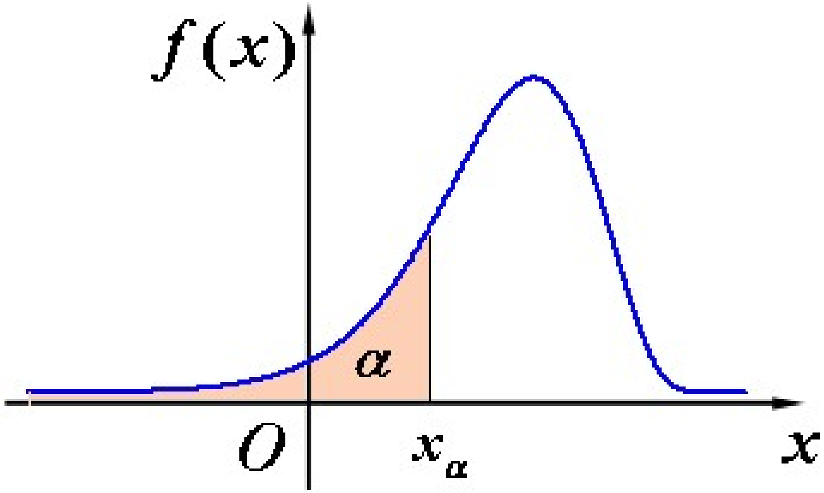
	总体	样本
均值	$\mu=\sum_{j=1}^{+\infty}p_j\cdot x_j$ $\downarrow\quad\downarrow$ 权重(概率) 总体X的理论值	$\bar{X}=\sum_{i=1}^{n}\frac{1}{n}\cdot X_i$ $\downarrow\quad\downarrow$ 权重(频率) 样本值
方差	$\sigma^2=\sum_{j=1}^{+\infty}p_j\cdot(x_j-\mu)^2$	$S^2=\sum_{i=1}^{n}\frac{1}{n}\cdot(X_i-\bar{X})^2$
l阶矩	$\mu_l=\sum_{j=1}^{+\infty}p_j\cdot x_j^l$	$A_l=\sum_{i=1}^{n}\frac{1}{n}\cdot X_i^l$
协方差	$Cov(X,Y)=\sum_{j=1}^{+\infty}\sum_{k=1}^{+\infty}p_{jk}\cdot(x_j-\mu_X)(y_k-u_Y)$	$S_{XY}=\sum_{i=1}^{n}\frac{1}{n}\cdot(X_i-\bar{X})(Y_i-\bar{Y})$
相关系数	$\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{DX}\sqrt{DY}}$	$\hat{\rho}_{XY}=\frac{S_{XY}}{\sqrt{S_{XX}}\sqrt{S_{YY}}}$
结果	实数	随机变量

二、统计分析中四大常用的分布

定义 1.4 设 $X \sim F(x)$, 若 x_α 满足

$$\boxed{F(x_\alpha) = P\{X < x_\alpha\} = \alpha, \quad 0 < \alpha < 1}$$

则称 x_α 为分布 $F(x)$ 的 α 分位数(点).



可见, $\boxed{x_\alpha = F^{-1}(\alpha)}$.

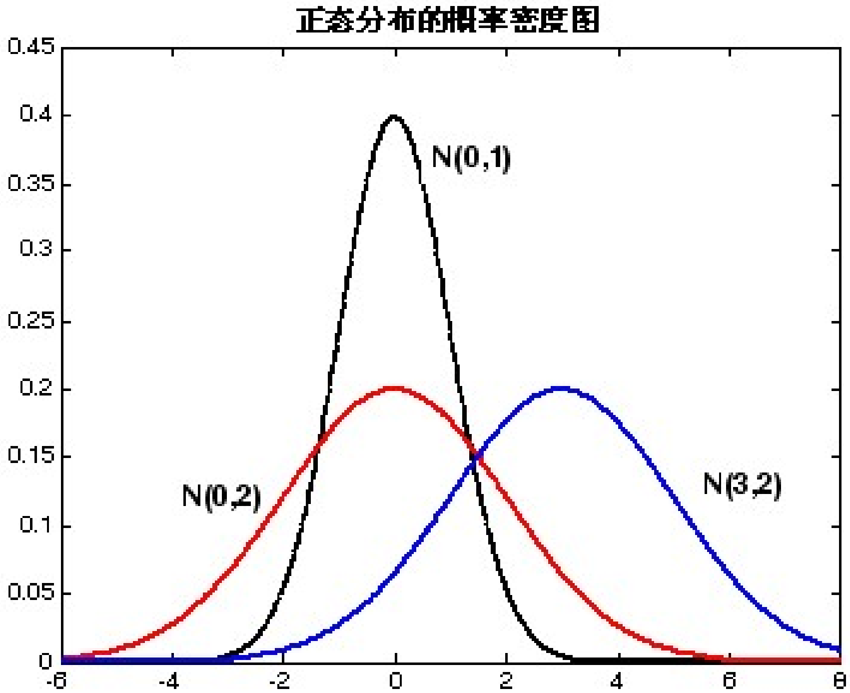
(一) $N(\mu,\sigma^2)$分布

1.定义

$N(\mu,\sigma^2)$分布的概率密度:

$$f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

$$-\infty<x<+\infty$$



2.可加性 (其中a,b为常数)

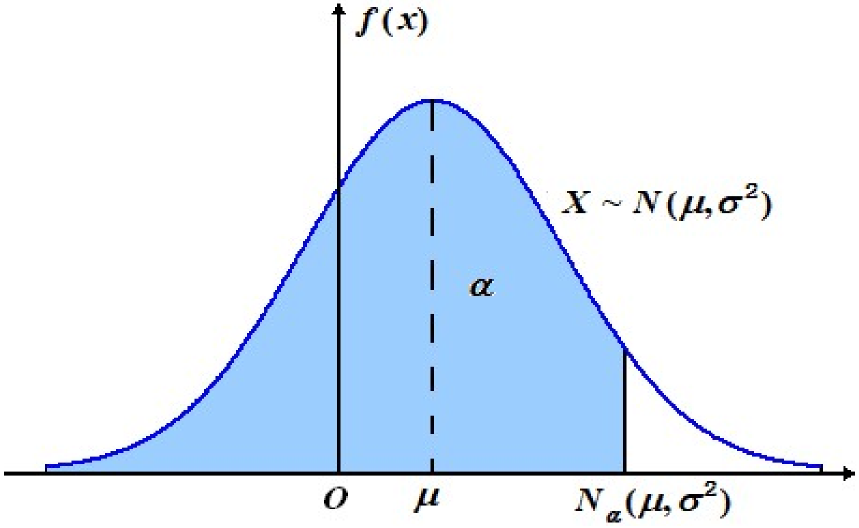
$$\left.\begin{array}{l}X\sim N(\mu_1,\sigma_1^2)\\ Y\sim N(\mu_2,\sigma_2^2)\end{array}\right\}\xrightarrow{X\text{与}Y\text{独立}}aX+bY\sim N(a\mu_1+b\mu_2,a^2\sigma_1^2+b^2\sigma_2^2)$$

3.期望和方差

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

4. 分位数 $N_{\alpha}(\mu,\sigma^2)$

$$P\{X < N_{\alpha}(\mu,\sigma^2)\} = \alpha$$

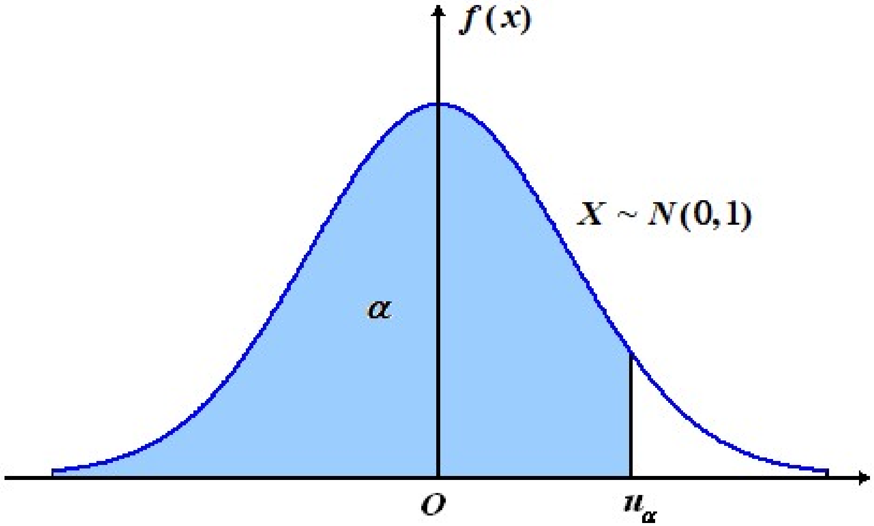


1-8(a) $N(\mu,\sigma^2)$的α分位数$N_{\alpha}(\mu,\sigma^2)$

$N(0,1)$的分位数u_{α}

$$P\{X < u_{\alpha}\} = \alpha$$

u_{α}性质: $\boxed{u_{1-\alpha} = -u_{\alpha}}$



$N(0,1)$的α分位数u_{α}

(二) χ^2分布

1.定义:

$$\chi^2 = \sum_{i=1}^{n} X_i^2 \quad \underset{X_i\text{相互独立}}{\overset{X_i \sim N(0,1)}{\sim\sim\sim\sim\sim\sim}} \quad \chi^2(n)$$

χ^2分布的概率密度:

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

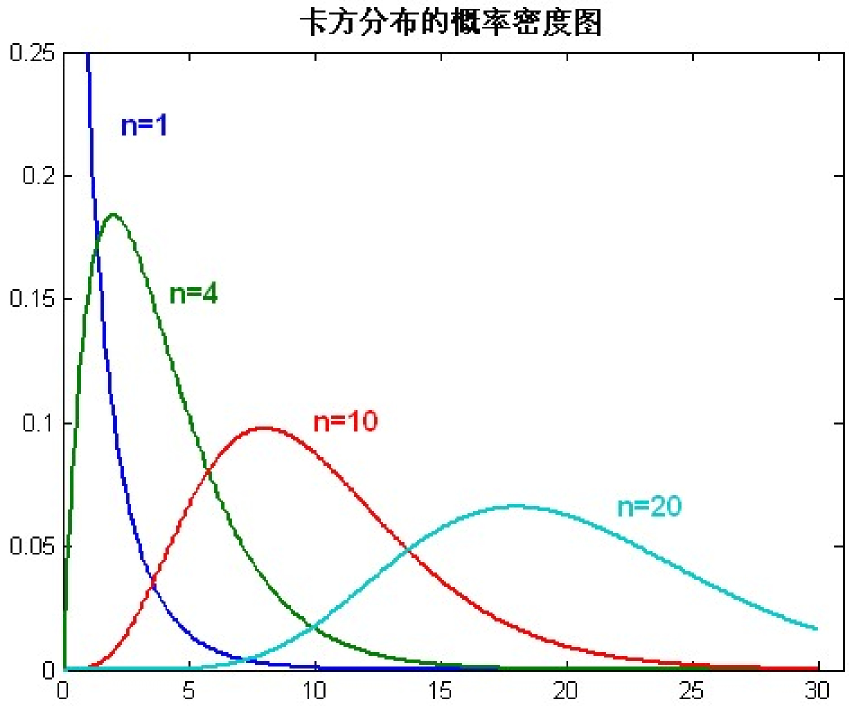


图 1-4

2.可加性:

$$\left.\begin{array}{l}\chi_1^2 \sim \chi^2(n_1)\\ \chi_2^2 \sim \chi^2(n_2)\end{array}\right\} \xrightarrow{\chi_1^2\text{与}\chi_2^2\text{独立}} \chi_1^2+\chi_2^2 \sim \chi^2(n_1+n_2).$$

3.期望和方差

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

4.分位数$\chi_\alpha^2(n)$

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

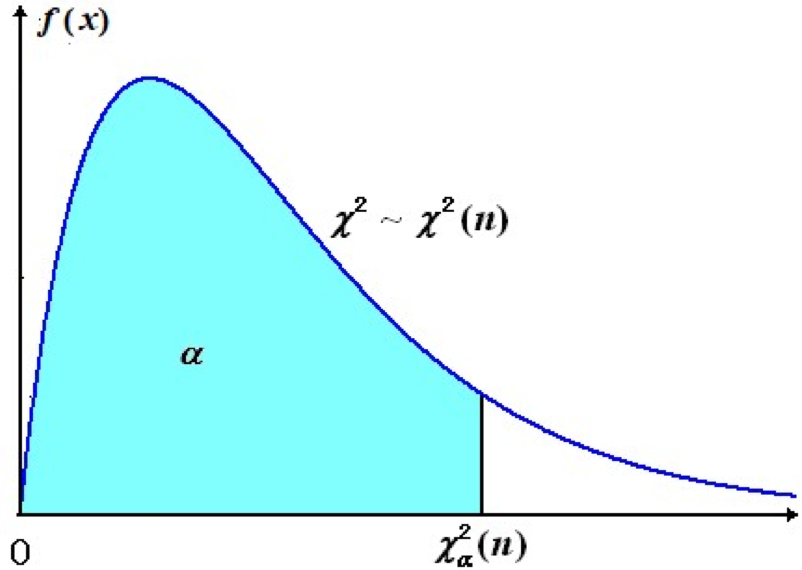


图 1-8(b) $\chi^2(n)$分布的α分位数$\chi_\alpha^2(n)$

(三) t分布

1.定义:

$$\left.\begin{array}{l} X \sim N(0,1) \\ Y \sim \chi^2(n) \end{array}\right\} \xrightarrow{X与Y独立} t=\frac{X}{\sqrt{Y/n}} \sim t(n)$$

t分布的概率密度:

$$f(x)=\frac{\Gamma[(n+1)/2]}{\sqrt{\pi n}\,\Gamma(n/2)}\left(1+\frac{x^2}{n}\right)^{-(n+1)/2},$$

$$-\infty < x < \infty$$

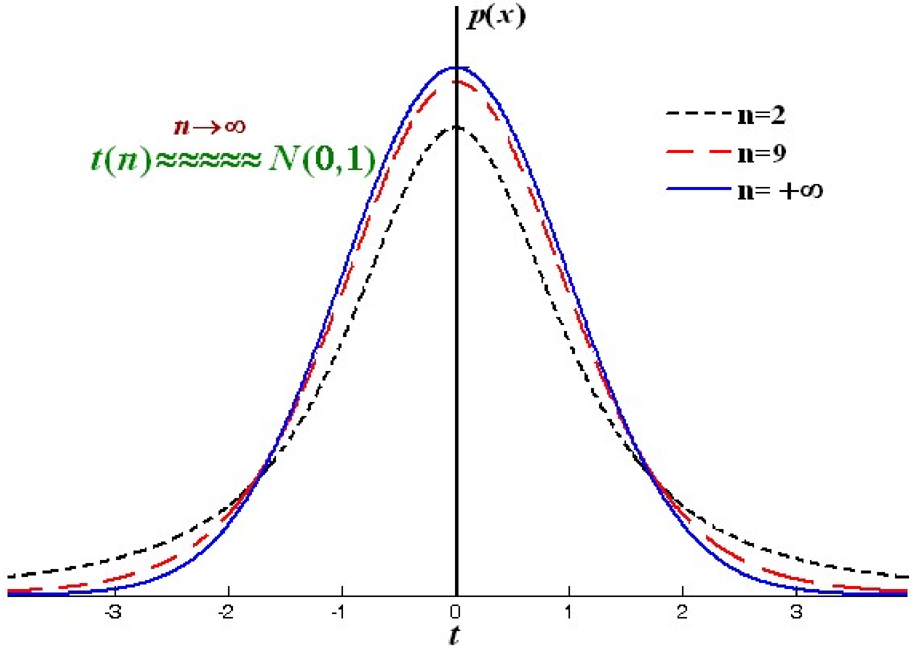


图1-5

2.期望与方差

$$E(t)=0(n>1)$$

$$D(t)=\frac{n}{n-2}(n>2)$$

3.分位数$t_\alpha(n)$

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

分位数性质: $\boxed{t_{1-\alpha}(n)=-t_\alpha(n)}$

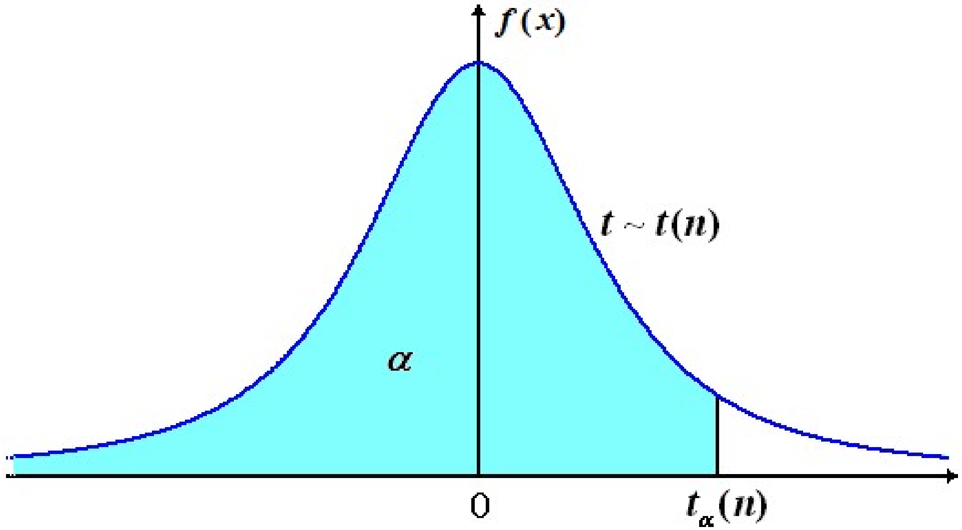


图1-8(c) $t(n)$分布的α分位数$t_\alpha(n)$

(四) F分布

1.定义:

$$\left.\begin{array}{l} U \sim \chi^2(n_1) \\ V \sim \chi^2(n_2) \end{array}\right\} \xrightarrow{U与V独立} F = \frac{U/n_1}{V/n_2} \sim F(n_1, n_2)$$

F分布的概率密度:

$$f(x) = \begin{cases} \dfrac{\Gamma[(n_1+n_2)/2]}{\Gamma(n_1/2)\Gamma(n_2/2)} n_1^{\frac{n_1}{2}} n_2^{\frac{n_2}{2}} \dfrac{x^{n_1/2-1}}{(n_1 x + n_2)^{(n_1+n_2)/2}}, & x > 0 \\ 0, & 其它 \end{cases}$$

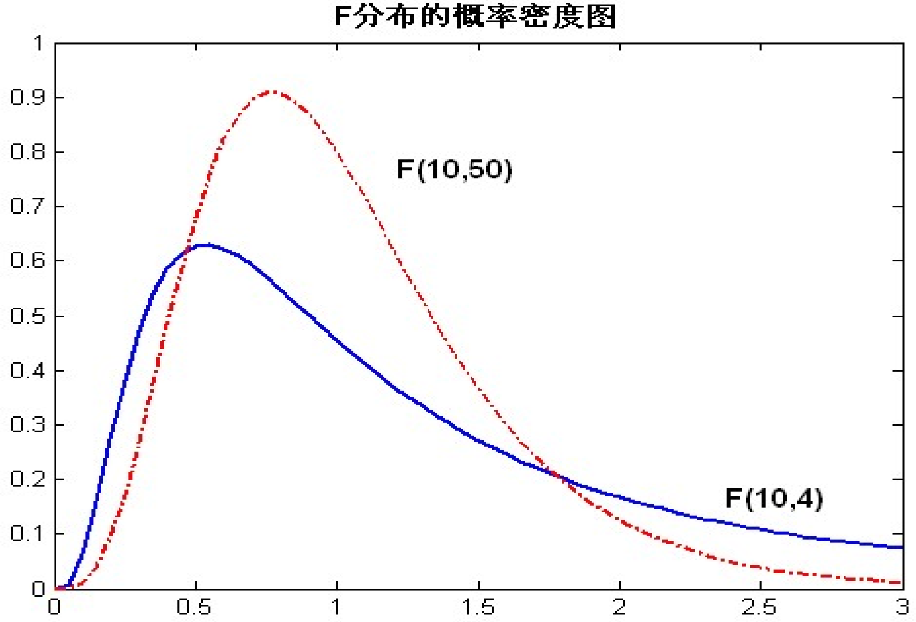


图1-6

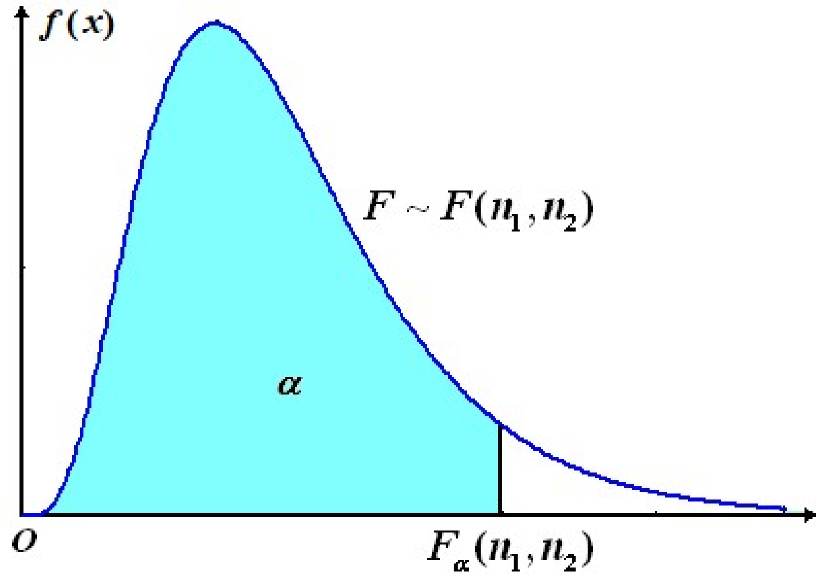


图1-8(d) F分布的α分位数$F_\alpha(n_1,n_2)$

2.分位数$F_\alpha(n_1,n_2)$

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

【★例1.13(P22)】求下列分位数:

(i) $u_{0.9}$; (ii) $t_{0.25}(4)$; (iii) $F_{0.1}(14,10)$; (iv) $\chi^2_{0.025}(50)$.

三、正态总体的抽样分布

1.(对任意分布的总体)

设$X_1,X_2,\cdots,X_n$是来自总体X的样本，且$E(X)=\mu$，$D(X)=\sigma^2$，则

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

$$\left.\begin{aligned}
&E(S^2)=E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\right]=\frac{n-1}{n}\sigma^2,\\
&E(S_*^2)=E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\right]=\sigma^2,\\
&E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right]=\sigma^2
\end{aligned}\right\}
\xRightarrow{X\sim N(\mu,\sigma^2)}
\left\{\begin{aligned}
&D(S^2)=\frac{2(n-1)}{n^2}\sigma^4,\\
&D(S_*^2)=\frac{2}{n-1}\sigma^4,\\
&D\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right]=\frac{2}{n}\sigma^4
\end{aligned}\right.$$

2.(对一个正态总体)

设总体 $X \sim N(\mu,\sigma^2)$, $X_1,X_2,\cdots,X_n$是来自X的样本, 记

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

(1) $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$;

(2) $\dfrac{\bar{X}-\mu}{S_*/\sqrt{n}} \sim t(n-1)$;

(3) $\dfrac{(n-1)S_*^2}{\sigma^2}=\dfrac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2} \sim \chi^2(n-1)$, 且$\bar{X}$与$S_*^2$独立;

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

3.(对两个正态总体)设

$X_1, X_2, \cdots, X_{n_1}$是来自总体$X \sim N(\mu_1, \sigma_1^2)$的样本,

$Y_1, Y_2, \cdots, Y_{n_2}$是来自总体$Y \sim N(\mu_2, \sigma_2^2)$的样本,

且这两个样本相互独立,则有

(1) $$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}} \sim N(0,1);$$

(2) 当$\sigma_1^2=\sigma_2^2=\sigma^2$时,记$S_w=\sqrt{\dfrac{(n_1-1)S_{1*}^2+(n_2-1)S_{2*}^2}{n_1+n_2-2}}$,则

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

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

(4)(补) $\dfrac{\dfrac{1}{n_1}\sum\limits_{i=1}^{n_1}(X_i - \mu_1)^2 \Big/ \dfrac{1}{n_2}\sum\limits_{j=1}^{n_2}(Y_j - \mu_2)^2}{\sigma_1^2 / \sigma_2^2} \sim F(n_1, n_2)$.

作业:

($P_{26\text{-}28}$) 1.1, 1.12