§6.2 抽样分布

确定统计量的分布是数理统计的基本问题之一

正态总体是最常见的总体，本节介绍的几个抽样分布均对正态总体而言.

统计中常用分布

(1) 正态分布

若 $X_1, X_2, \cdots, X_n \sim N(\mu_i, \sigma_i^2)$

则 $$\sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

特别地,

若 $X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} X_i \sim N(\mu, \sigma^2)$

则 $$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

标准正态分布的 α 分位数

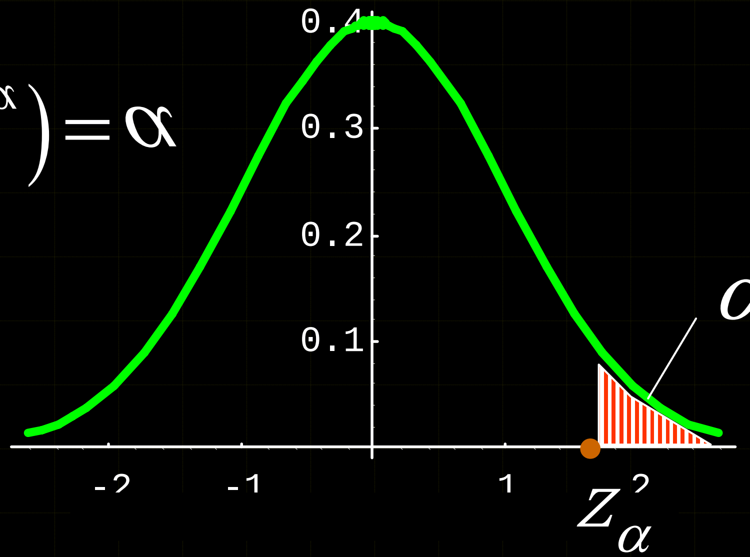
定义

若 $P(X > z_\alpha) = \alpha$，则称z_α为标准正态分布的上α 分位数.

若 $P(|X| > z_{\frac{\alpha}{2}}) = \alpha$，则称 $z_{\frac{\alpha}{2}}$ 为标准正态分布的双侧 α 分位数.

标准正态分布的α 分位数图形

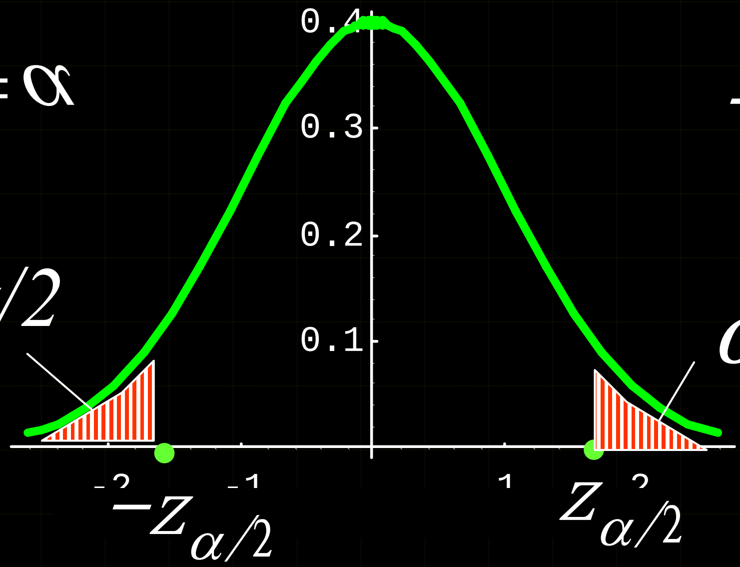
$$P(X > z_\alpha) = \alpha$$



$$\left.\begin{aligned} z_{0.05} &= 1.645 \\ z_{0.025} &= 1.96 \end{aligned}\right\}\ \boxed{常用数字}$$

$$z_{0.005} = 2.575$$

$$P(|X| > z_{\frac{\alpha}{2}}) = \alpha$$



$$-z_{\alpha/2} = z_{1-\alpha/2}$$

(2) $\chi^2(n)$分布（ n为自由度 ）

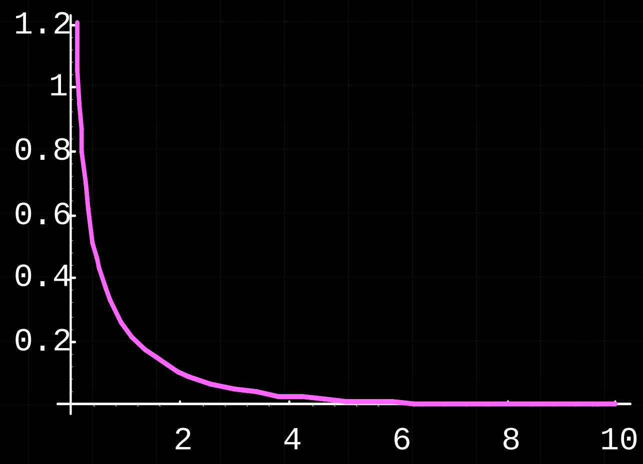
定义 设 $X_1, X_2, \cdots, X_n$相互独立，

且都服从标准正态分布$N(0, 1)$，则

$$\sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$$

$n=1$ 时，其密度函数为

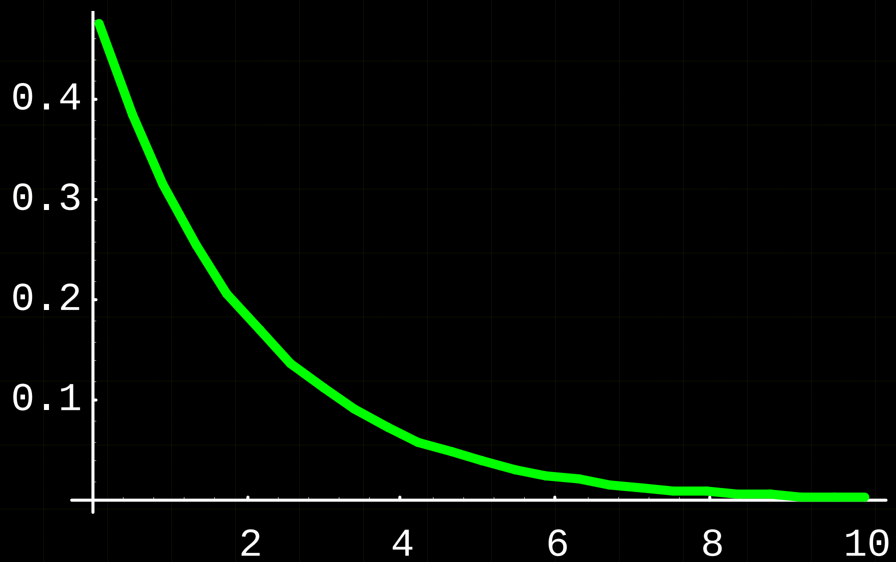
$$f(x) = \begin{cases} \dfrac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{x}{2}}, & x > 0 \\ 0, & x \le 0 \end{cases}$$



$n = 2$ 时，其密度函数为

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

为参数为1/2的指数分布.



一般 自由度为 n 的 $\chi^2(n)$ 的密度函数为

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

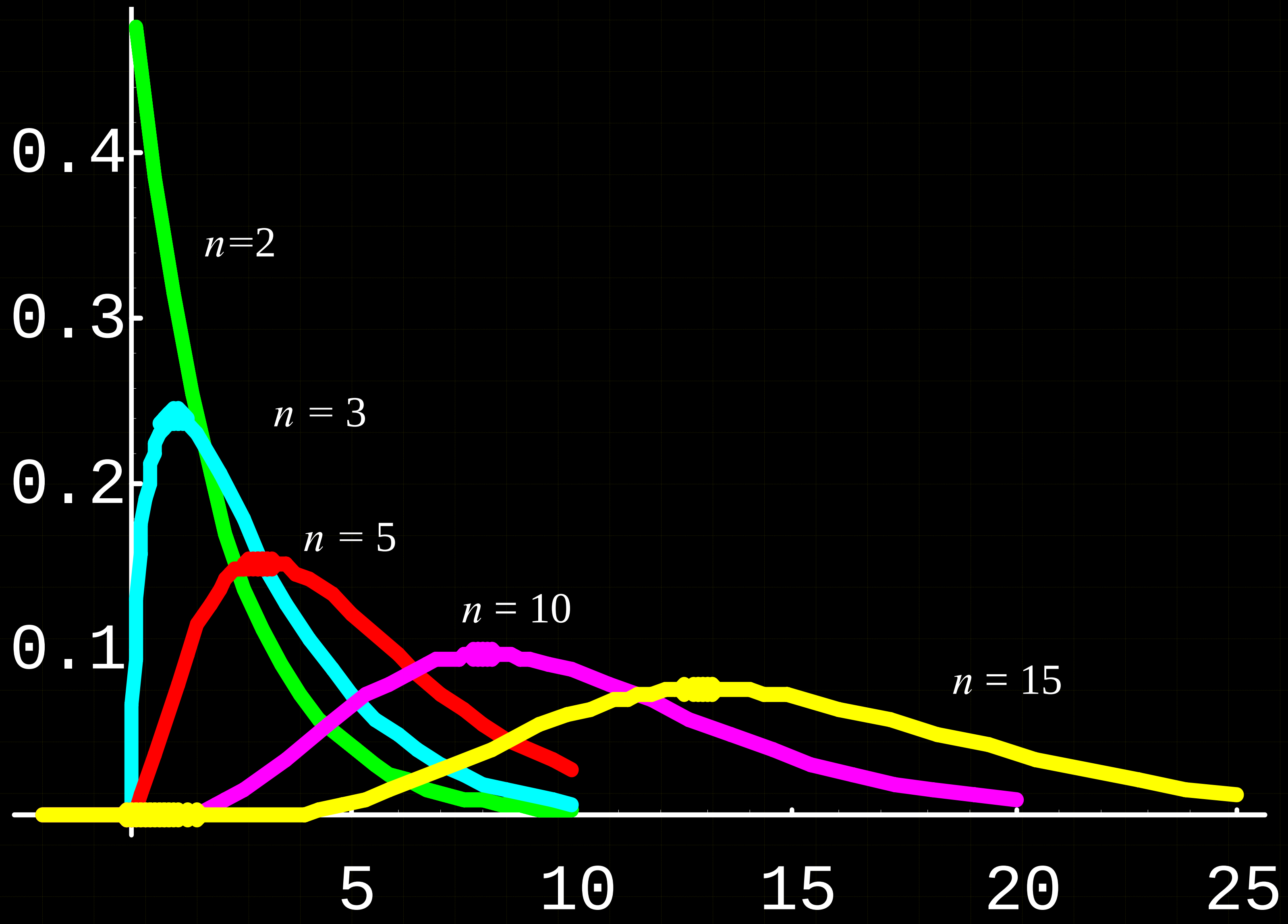
其中，

$$\Gamma(x)=\int_0^{+\infty}t^{x-1}e^{-t}dt$$

在$x>0$时收敛，称为Γ函数，具有性质

$$\Gamma(x+1)=x\Gamma(x),\quad \Gamma(1)=1,\ \Gamma(1/2)=\sqrt{\pi}$$

$$\Gamma(n+1)=n!\ (n\in N)$$


n=2
n = 3
n = 5
n = 10
n = 15
0.4
0.3
0.2
0.1
5
10
15
20
25

$\chi^2(n)$ 分布的性质

1° $E\left(\chi^2(n)\right) = n,\ D\left(\chi^2(n)\right) = 2n$

2° 若$X_1 = \chi^2(n_1), X_2 = \chi^2(n_2), X_1, X_2$相互独立

则$X_1 + X_2 \sim \chi^2(n_1 + n_2)$

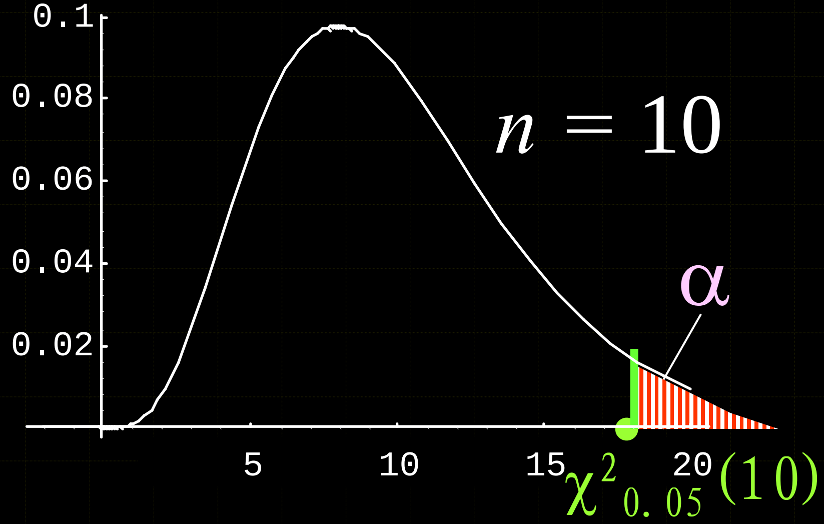
3° $n \to \infty$时，$\chi^2(n) \to$正态分布

4° $\chi^2(n)$分布的上 α 分位数有表可查

例如

$P\left(\chi^2(10) > 18.307\right) = 0.05$

$\chi^2_{0.05}(10) = 18.307$



证 1°设 $\chi^2(n)=\sum_{i=1}^{n}X_i^2 \quad X_i \sim N(0,1) \;\; i=1,2,\cdots,n$

$X_1, X_2, \cdots, X_n$ 相互独立，

则 $E(X_i)=0,\; D(X_i)=1,\; E(X_i^2)=1$

$$E\left(\chi^2(n)\right)=E\left(\sum_{i=1}^{n}X_i^2\right)=n$$

$$E(X_i^4)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}x^4 e^{-\frac{x^2}{2}}\mathrm{d}x=3$$

$$D(X_i^2)=E(X_i^4)-E^2(X_i^2)=2$$

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

(3) t 分布 (Student 分布)

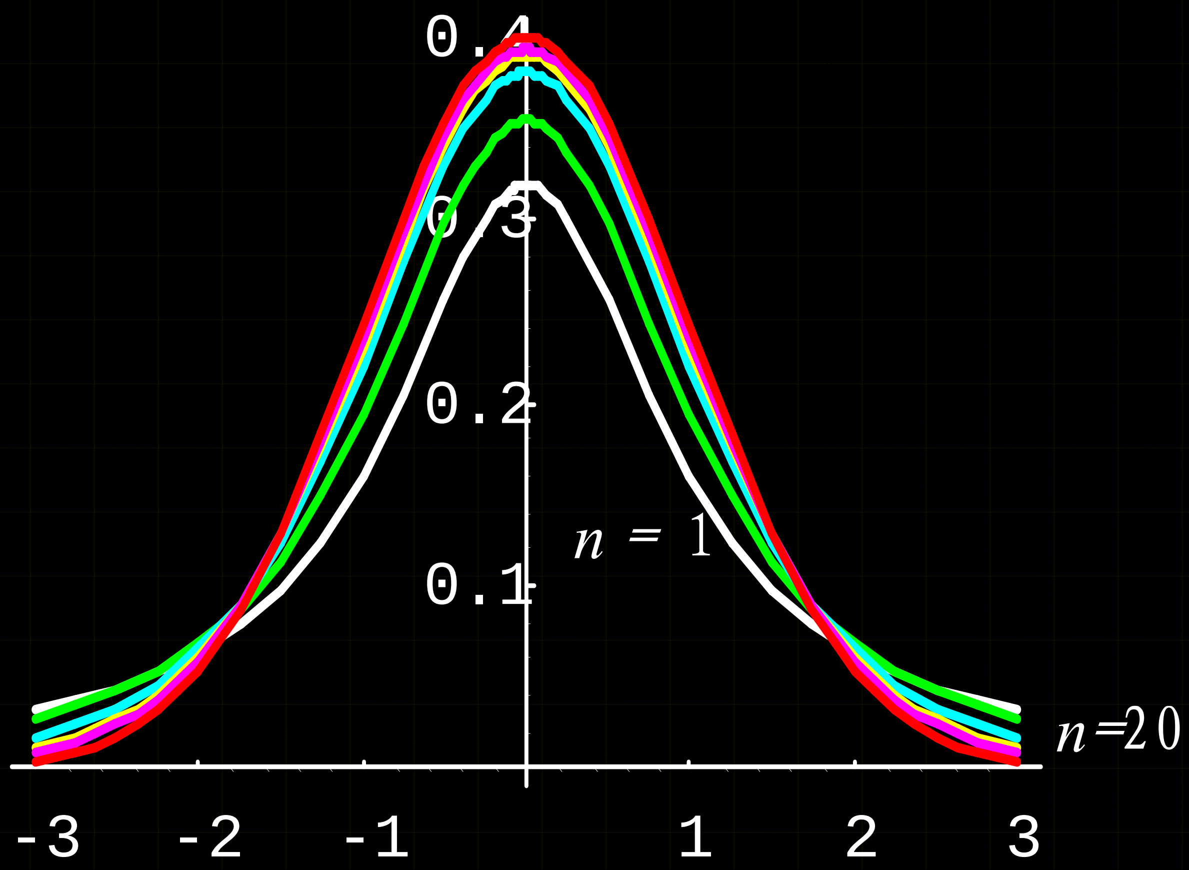
定义 设 $X \sim N(0,1)$，$Y \sim \chi^2(n)$, X, Y相互独立,

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

则称 T 服从自由度为 n 的T分布.

其密度函数为

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



t 分布的图形(红色的是标准正态分布)

t 分布的性质

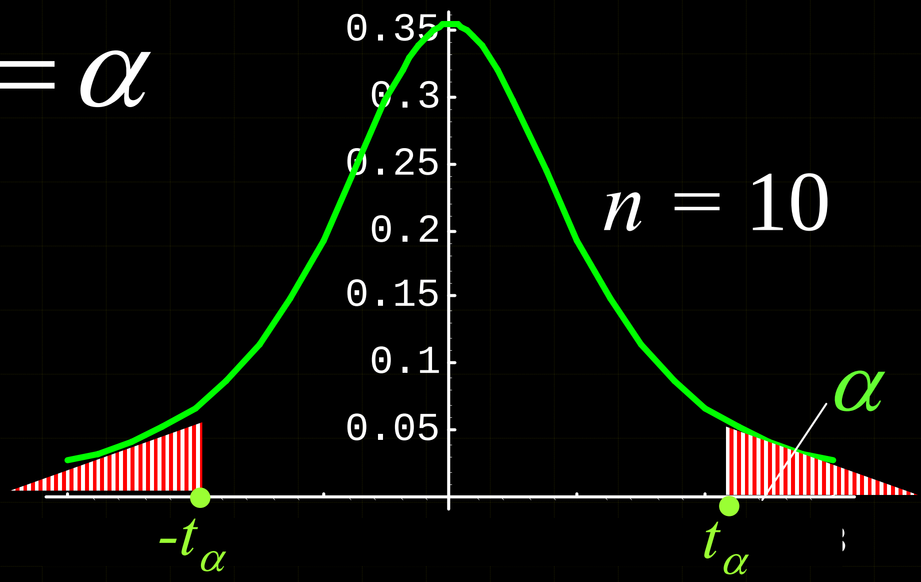
1° $f_n(t)$是偶函数,

$$n \to \infty, f_n(t) \to \varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

2° T分布的上α分位数 t_α与双测 α 分位数 $t_{\alpha/2}$ 均有表可查.

$$P(T > t_\alpha) = \alpha$$

$$-t_\alpha = t_{1-\alpha}$$



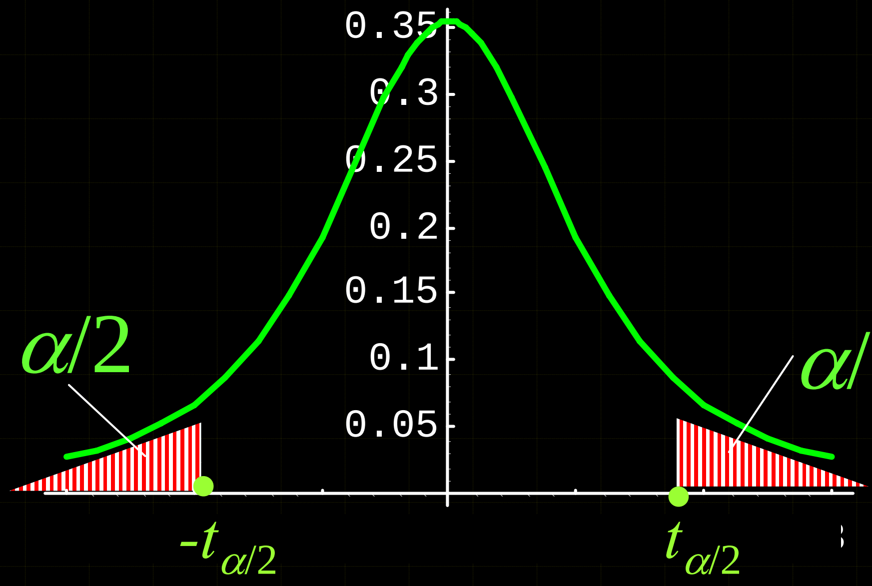
$$P(T > 1.8125) = 0.05 \Rightarrow t_{0.05}(10) = 1.8125$$

$$P(T < -1.8125) = 0.05, \quad P(T > -1.8125) = 0.95$$

$$\Rightarrow t_{0.95}(10) = -1.8125$$

$$P(T > t_{\alpha/2}) = \frac{\alpha}{2}$$

$$P(|T| > t_{\alpha/2}) = \alpha$$



$$P(T > 2.2281) = 0.025$$

$$P(|T| > 2.2281) = 0.05$$

$$\Rightarrow t_{0.025}(10) = 2.2281$$

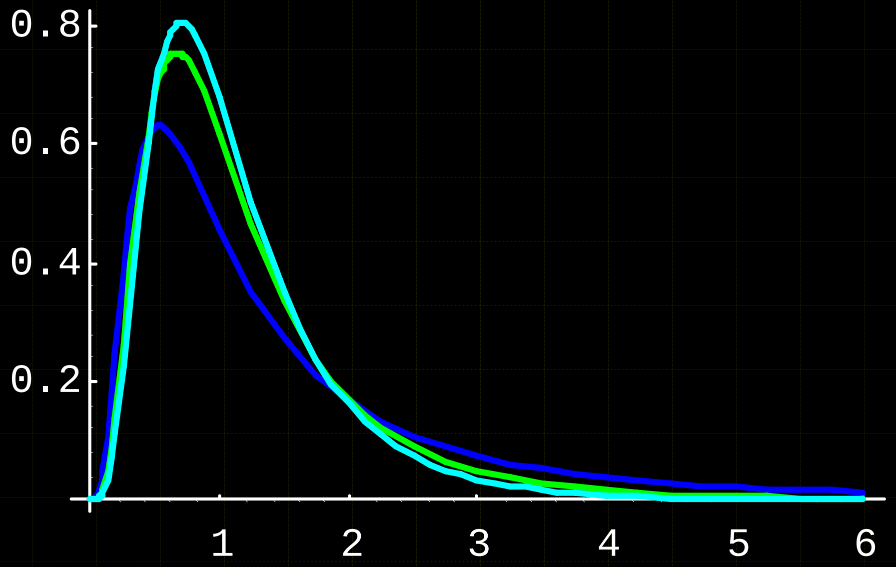
(4) F 分布

定义 设 $X\sim\chi^2(n), Y\sim\chi^2(m), X, Y$ 相互独立，

令
$$F=\frac{X/n}{Y/m}$$

则称 F 服从为**第一自由度为n，第二自由度为 m 的F 分布.** 其密度函数为

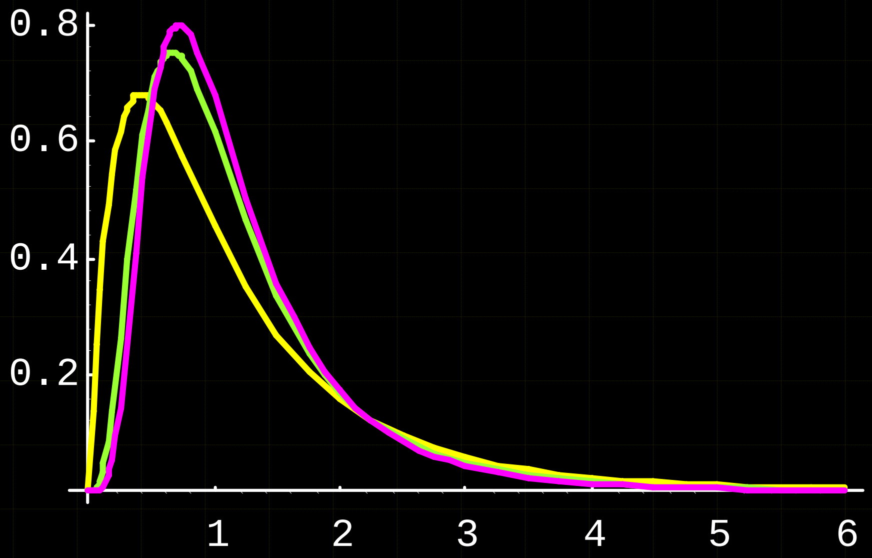
$$f(t,n,m)=\begin{cases}\dfrac{\Gamma\left(\dfrac{n+m}{2}\right)}{\Gamma\left(\dfrac{n}{2}\right)\Gamma\left(\dfrac{m}{2}\right)}\left(\dfrac{n}{m}\right)^{\frac{n}{2}}t^{\frac{n}{2}-1}\left(1+\dfrac{n}{m}t\right)^{-\frac{n+m}{2}} & t>0\\ 0, & t\le 0\end{cases}$$



$m = 10, n = 4$

$m = 10, n = 10$

$m = 10, n = 15$



$m = 4, n =10$

$m = 10, n = 10$

$m = 15, n = 10$

F 分布的性质

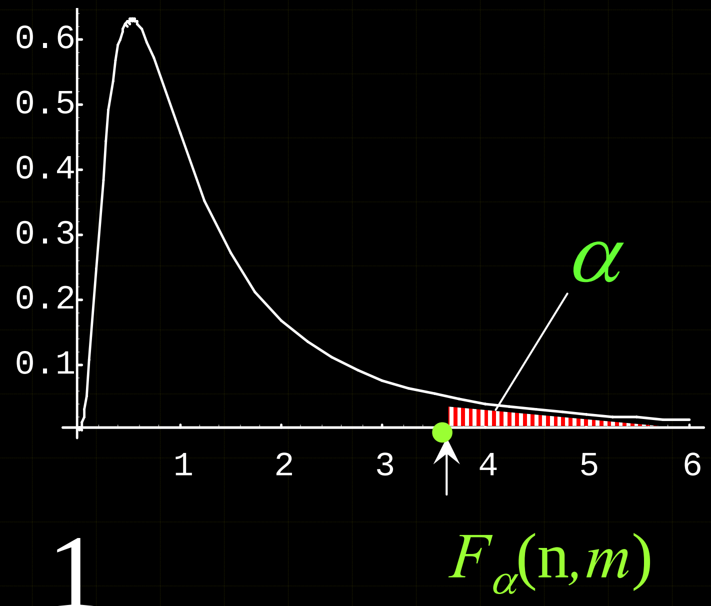
1° 若 $F \sim F(n,m)$，则 $1/F \sim F(m,n)$

2° $F(n,m)$ 的上 α 分位数 $F_\alpha(n,m)$ 有表可查:

$$P(F > F_\alpha(n,m)) = \alpha$$

例如 $F_{0.05}(4,5) = 5.19$

求 $F_{0.95}(5,4) = ?$



事实上，$F_{1-\alpha}(n,m) = \dfrac{1}{F_\alpha(m,n)}$

故 $F_{0.95}(5,4) = \dfrac{1}{F_{0.05}(4,5)} = \dfrac{1}{5.19}$

例1 证明 $F_{1-\alpha}(n,m)=\dfrac{1}{F_{\alpha}(m,n)}$

证 $P(F \geq F_{1-\alpha}(n,m)) = P\left(\dfrac{1}{F} \leq \dfrac{1}{F_{1-\alpha}(n,m)}\right)$

$$=1-P\left(\frac{1}{F} \geq \frac{1}{F_{1-\alpha}(n,m)}\right)=1-\alpha$$

故 $P\left(\dfrac{1}{F} \geq \dfrac{1}{F_{1-\alpha}(n,m)}\right)=\alpha$ 由于 $\dfrac{1}{F} \sim F(m,n)$

因而 $\dfrac{1}{F_{1-\alpha}(n,m)}=F_{\alpha}(m,n)$

例2 证明： $[t_{1-\frac{\alpha}{2}}(n)]^2 = F_\alpha(1,n)$

证 设 $X \sim T(n)$, $X = G\Big/\sqrt{\dfrac{\chi^2(n)}{n}}$, $G \sim N(0,1)$

令 $$Y = X^2 = \frac{G^2}{\dfrac{\chi^2(n)}{n}} = \frac{\dfrac{\chi^2(1)}{1}}{\dfrac{\chi^2(n)}{n}} \sim F(1,n)$$

有 $$P(|X| > |t_{1-\frac{\alpha}{2}}(n)|) = P(|X| > t_{\frac{\alpha}{2}}(n)) = \alpha$$

$$= P(X^2 > t^2_{\frac{\alpha}{2}}(n)) = P(Y > t^2_{1-\frac{\alpha}{2}}(n))$$

即 $$t^2_{1-\frac{\alpha}{2}}(n) = F_\alpha(1,n)$$

抽样分布的某些结论

（Ⅰ）一个正态总体

设总体 $X \sim N(\mu,\sigma^2)$，样本为（$X_1,\cdots,X_n$），

$$\overline{X} \sim N\left(\mu,\frac{\sigma^2}{n}\right) \Rightarrow \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

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

$\dfrac{(n-1)S^2}{\sigma^2}$ 与 $\overline{X}$ 相互独立 (1)

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

（II）两个正态总体

设 $X_1, X_2, \cdots, X_n$ 与 $Y_1, Y_2, \cdots, Y_m$ 分别是来自正态总体 $X \sim N(\mu_1, \sigma_1^2)$ 与 $Y \sim N(\mu_2, \sigma_2^2)$ 的相互独立的简单随机样本.

令 $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ $\qquad \overline{Y} = \dfrac{1}{m}\sum_{j=1}^{m} Y_j$

$$S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \qquad S_2^2 = \frac{1}{m-1}\sum_{j=1}^{m}(Y_j - \overline{Y})^2$$

则 $\dfrac{(n-1)S_1^2}{\sigma_1^2} \sim \chi^2(n-1)$ $\dfrac{(m-1)S_2^2}{\sigma_2^2} \sim \chi^2(m-1)$

$$\boxed{\frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \sim F(n-1, m-1)} \quad \cdots\cdots \quad \textbf{(3)}$$

若 $\sigma_1 = \sigma_2$ 则 $\dfrac{S_1^2}{S_2^2} \sim F(n-1, m-1)$

设 $X_1, X_2, \cdots, X_n$ 与 $Y_1, Y_2, \cdots, Y_m$ 分别是来自正态总体 $X \sim N(\mu_1, \boxed{\sigma^2})$ 与 $Y \sim N(\mu_2, \boxed{\sigma^2})$ 的相互独立的简单随机样本.

则 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu_1, \frac{\sigma^2}{n})$ $\quad \overline{Y} = \frac{1}{m}\sum_{j=1}^{m} Y_j \sim N(\mu_2, \frac{\sigma^2}{m})$

$$\longrightarrow \quad \overline{X} - \overline{Y} \sim N(\mu_1 - \mu_2, \frac{\sigma^2}{n} + \frac{\sigma^2}{m})$$

$$\longrightarrow \quad \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}} \sim N(0,1)$$

$$\frac{(n-1)S_1^2}{\sigma^2} \sim \chi^2(n-1) \quad \frac{(m-1)S_2^2}{\sigma^2} \sim \chi^2(m-1)$$

$$\Longrightarrow \frac{(n-1)S_1^2}{\sigma^2} + \frac{(m-1)S_2^2}{\sigma^2} \sim \chi^2(n+m-2)$$

$\overline{X} - \overline{Y}$ 与 $\dfrac{(n-1)S_1^2}{\sigma^2} + \dfrac{(m-1)S_2^2}{\sigma^2}$ 相互独立

$$\frac{\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma^2}{n}+\dfrac{\sigma^2}{m}}}}{\sqrt{\dfrac{\dfrac{(n-1)S_1^2}{\sigma^2}+\dfrac{(m-1)S_2^2}{\sigma^2}}{n+m-2}}}$$

$$=\boxed{\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}\sqrt{\dfrac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2}}}\sim t(n+m-2)}$$

(4)

例3 设 $X \sim N(72, 100)$, 为使样本均值大于70的概率不小于90%, 则样本容量至少取多少?

解 设样本容量为 n, 则 $\overline{X} \sim N(72, \frac{100}{n})$

故 $P(\overline{X} > 70) = 1 - P(\overline{X} \le 70)$

$$= 1 - \Phi\left(\frac{70 - 72}{\frac{10}{\sqrt{n}}}\right) = \Phi\left(0.2\sqrt{n}\right)$$

令 $\Phi\left(0.2\sqrt{n}\right) \ge 0.9$ 得 $0.2\sqrt{n} \ge 1.29$

即 $n \ge 41.6025$ 所以取 $n = 42$

例4 从正态总体 $X \sim N(\mu, \sigma^2)$ 中，抽取了 $n = 20$的样本$(X_1, X_2, \cdots, X_{20})$

(1) 求$P\left(0.37\sigma^2 \le \frac{1}{20}\sum_{i=1}^{20}(X_i - \overline{X})^2 \le 1.76\sigma^2\right)$

(2) 求$P\left(0.37\sigma^2 \le \frac{1}{20}\sum_{i=1}^{20}(X_i - \mu)^2 \le 1.76\sigma^2\right)$

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

即 $\frac{19S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{20}\left(X_i - \overline{X}\right)^2 \sim \chi^2(19)$

故

$$P\left(0.37\sigma^2 \le \frac{1}{20}\sum_{i=1}^{20}\left(X_i - \overline{X}\right)^2 \le 1.76\sigma^2\right)$$

$$= P\left(7.4 \le \frac{1}{\sigma^2}\sum_{i=1}^{20}\left(X_i - \overline{X}\right)^2 \le 35.2\right)$$

$$= P\left(\frac{1}{\sigma^2}\sum_{i=1}^{20}\left(X_i - \overline{X}\right)^2 \ge 7.4\right) - P\left(\frac{1}{\sigma^2}\sum_{i=1}^{20}\left(X_i - \overline{X}\right)^2 \ge 35.2\right)$$

查表

$$= 0.99 - 0.01 = 0.98$$

(P.386)

(2) $$\sum_{i=1}^{20}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(20)$$

故 $$P\left(0.37\sigma^2 \le \frac{1}{20}\sum_{i=1}^{20}(X_i-\mu)^2 \le 1.76\sigma^2\right)$$

$$=P\left(7.4 \le \sum_{i=1}^{20}\left(\frac{X_i-\mu}{\sigma}\right)^2 \le 35.2\right)$$

$$=P\left(\sum_{i=1}^{20}\left(\frac{X_i-\mu}{\sigma}\right)^2 \ge 7.4\right)-P\left(\sum_{i=1}^{20}\left(\frac{X_i-\mu}{\sigma}\right)^2 \ge 35.2\right)$$

$$=0.995-0.025=0.97$$

例5 设r.v. X 与 Y 相互独立，$X \sim N(0,16)$, $Y \sim N(0,9)$, $X_1, X_2, \ldots, X_9$ 与 $Y_1, Y_2, \ldots, Y_{16}$ 分别是取自 X 与 Y 的简单随机样本, 求统计量

$$Z = \frac{X_1 + X_2 + \cdots + X_9}{\sqrt{Y_1^2 + Y_2^2 + \cdots + Y_{16}^2}}$$

所服从的分布.

解 $X_1 + X_2 + \cdots + X_9 \sim N(0, 9 \times 16)$

$$\frac{1}{3 \times 4}(X_1 + X_2 + \cdots + X_9) \sim N(0,1)$$

$$\frac{1}{3}Y_i \sim N(0,1) \quad ,i = 1,2,\cdots,16$$

$$\sum_{i=1}^{16}\left(\frac{1}{3}Y_i\right)^2 \sim \chi^2(16)$$

从而 $$\frac{X_1 + X_2 + \cdots + X_9}{\sqrt{Y_1^2 + Y_2^2 + \cdots + Y_{16}^2}}$$

$$= \frac{\dfrac{1}{3\times 4}\left(X_1 + X_2 + \cdots + X_9\right)}{\sqrt{\dfrac{\sum_{i=1}^{16}\left(\frac{1}{3}Y_i\right)^2}{16}}} \sim t(16)$$

例6 设总体 $X \sim N(0,1)$，$X_1,\cdots,X_6$ 为总体 X 的样本，$Y=(X_1+X_2+X_3)^2+(X_4+X_5+X_6)^2$ 试确定常数 c，使 cY 服从 χ^2 分布.

解 $X_1+X_2+X_3 \sim N(0,3)$，$X_4+X_5+X_6 \sim N(0,3)$

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

故 $$\left[\frac{1}{\sqrt{3}}\left(X_1+X_2+X_3\right)\right]^2+\left[\frac{1}{\sqrt{3}}\left(X_4+X_5+X_6\right)\right]^2$$

$$=\frac{1}{3}Y \sim \chi^2(2)$$

因此 $c=1/3$.

例7 设$(X_1, X_2, \cdots, X_n)$是来自$N(\mu, \sigma^2)$的简单随机样本，$\bar{X}$是样本均值，

$$S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2, \qquad S_2^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2,$$

$$S_3^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \mu)^2, \qquad S_4^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2,$$

则服从自由度为$n-1$的t分布的随机变量为

(A) $\dfrac{\bar{X}-\mu}{S_1}\sqrt{n-1}$ (B) $\dfrac{\bar{X}-\mu}{S_2}\sqrt{n-1}$

(C) $\dfrac{\bar{X}-\mu}{S_3}\sqrt{n}$ (D) $\dfrac{\bar{X}-\mu}{S_4}\sqrt{n}$

解 $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ $\quad \dfrac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2 \sim \chi^2(n-1)$

$$\frac{\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{\dfrac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}}} = \frac{\sqrt{n(n-1)}(\bar{X}-\mu)}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}} \sim t(n-1)$$

故应选 (B)

作业 P. 202 习题六

9　10

补充作业

1. 设$X_1, X_2, \cdots, X_{2n}\ (n \ge 2)$为从正态总体 $X \sim N(\mu, \sigma^2)$ 中抽取的简单随机样本

其样本均值为 $\overline{X} = \dfrac{1}{2n}\sum_{i=1}^{2n} X_i$

求统计量 $Y = \sum_{i=1}^{n}(X_i + X_{n+i} - 2\overline{X})^2$

的数学期望 $E(Y)$. $(\sigma > 0)$ ([redacted])

2. $\overline{X}, \overline{Y}$ 是来自正态总体 $N(\mu, \sigma^2)$ 的容量为 n 的两个样本均值, 且两样本相互独立, 试确定 n, 使两样本均值之差的绝对值超过 σ 的概率大约为 0.01.

第十三周 问 题

如何估计湖中黑、白鱼的比例

某水产养殖场两年前在人工湖中混养了黑、白两种鱼. 现在需要对黑白鱼数目的比例进行估计.

提示：分别用矩法与极大似然估计法解决此问题.