常用统计量的分布（抽样分布）

——来自正态总体的统计量

1. χ^2分布

2. t分布

3. F分布

1. χ^2分布

设$X_1,\dots,X_n$为来自标准正态总体$N(0,1)$的样本，

则称统计量：

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

为服从自由度为n的χ^2分布

记为 $\chi^2 \sim \chi^2(n)$

表示一个随机变量

(1)自由度为n

(2)每个$X_i \sim N(0,1)$

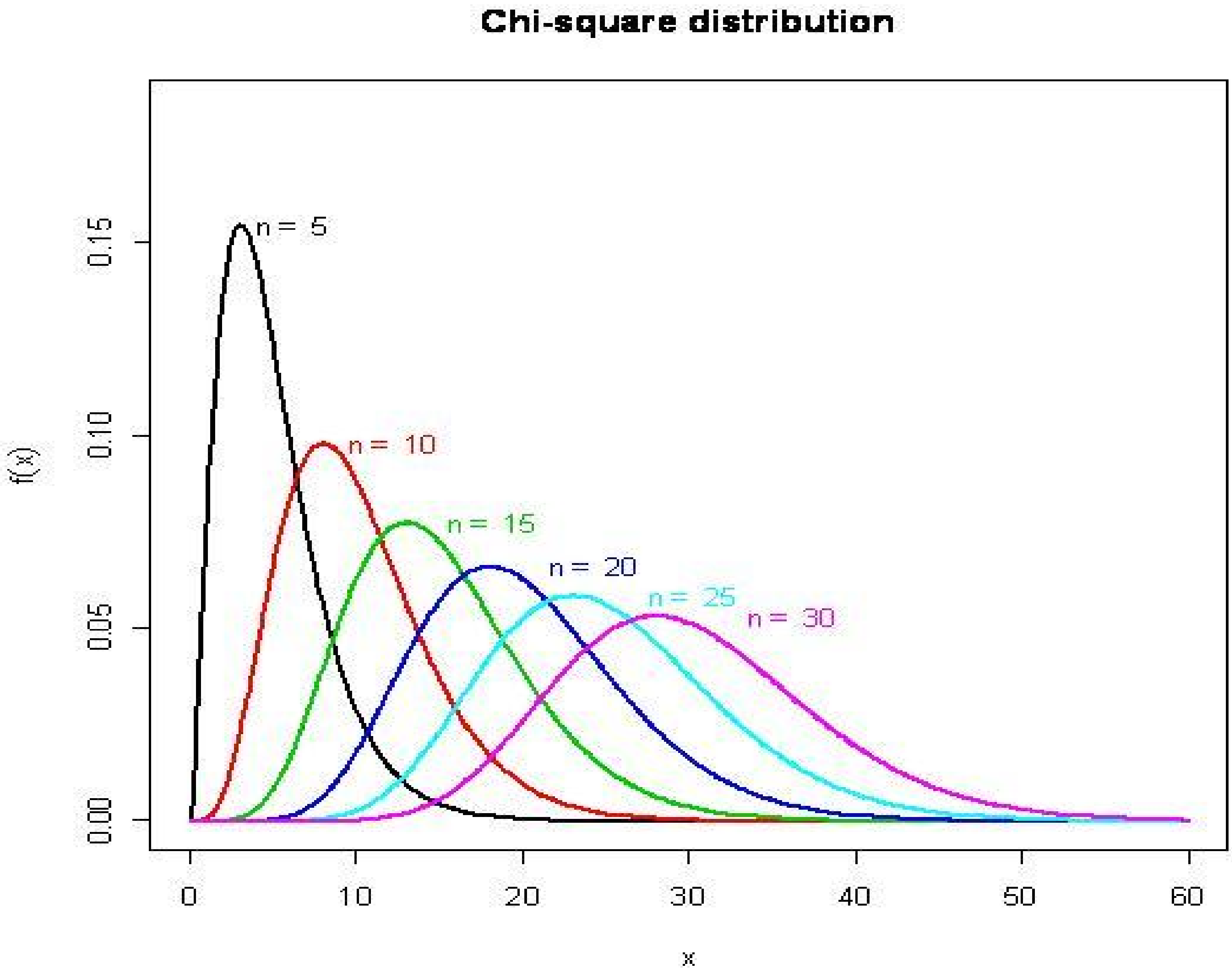
显然$X_i^2 \sim \chi^2(1)$.

密度函数

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

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

$\Gamma(n+1)=n!, \Gamma\left(\dfrac{1}{2}\right)=\sqrt{\pi}$


Chi-square distribution
n = 5
n = 10
n = 15
n = 20
n = 25
n = 30
f(x)
0.00
0.05
0.10
0.15
0
10
20
30
40
50
60
x

χ^2分布的性质

性质6.4.1 若$\chi^2 \sim \chi^2(n)$, 则$E(\chi^2) = n$, $D(\chi^2) = 2n$.

证 $\because \chi^2 \sim \chi^2(n)$

$\therefore \chi^2 = \sum_{i=1}^{n} X_i^2$并且$X_i \sim N(0,1)$

$$E(\chi^2) = E\left(\sum_{i=1}^{n} X_i^2\right) = \sum_{i=1}^{n} \underline{E(X_i^2)} = \sum_{i=1}^{n}\left[D(X_i) + (EX_i)^2\right] = \sum_{i=1}^{n}\left[1 + 0^2\right] = n,$$

$$E(X_i^2) = D(X_i) + \left(EX_i\right)^2$$

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2},-\infty<x<+\infty$$

X_i相互独立

$$D(X)=D\left(\sum_{i=1}^{n}X_i^{\,2}\right)=\sum_{i=1}^{n}\underline{D(X_i^{\,2})}=\sum_{i=1}^{n}\left[\underline{E(X_i^4)-(EX_i^{\,2})^2}\right]=\sum_{i=1}^{n}\left[3-1^2\right]=2n.$$

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

$$=-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}x^3d(e^{-\frac{x^2}{2}})=\underline{-\frac{1}{\sqrt{2\pi}}\left[x^3e^{-\frac{x^2}{2}}\Big|_{-\infty}^{+\infty}-3\int_{-\infty}^{+\infty}x^2e^{-\frac{x^2}{2}}dx\right]}$$

$$=\frac{3}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}x^2e^{-\frac{x^2}{2}}dx=3\int_{-\infty}^{+\infty}x^2\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=3E\left(X_i^{\,2}\right)=3.$$

分部积分.

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2},-\infty<x<+\infty$$

$$DX=D\left(\sum_{i=1}^{n}X_i^2\right)=\sum_{i=1}^{n}DX_i^2=\sum_{i=1}^{n}\left[EX_i^4-(EX_i^2)^2\right]=\sum_{i=1}^{n}\left[3-1^2\right]=2n.$$

$$EX_i^4=\int_{-\infty}^{+\infty}x^4\varphi(x)dx=2\int_0^{+\infty}x^4\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=\frac{2}{\sqrt{2\pi}}\int_0^{+\infty}x^3e^{-\frac{x^2}{2}}d\left(\frac{x^2}{2}\right)$$

$$\overset{令t=\frac{x^2}{2}}{=}\frac{2}{\sqrt{2\pi}}\int_0^{+\infty}(2t)^{3/2}e^{-t}dt=\frac{2}{\sqrt{2\pi}}\times2^{3/2}\times\int_0^{+\infty}t^{3/2}e^{-t}dt=\frac{4}{\sqrt{\pi}}\times\int_0^{+\infty}t^{\frac{5}{2}-1}e^{-t}dt$$

$$=\frac{4}{\sqrt{\pi}}\times\Gamma(\frac{5}{2})=\frac{4}{\sqrt{\pi}}\times\frac{3}{2}\times\Gamma(\frac{3}{2})=\frac{4}{\sqrt{\pi}}\times\frac{3}{2}\times\frac{1}{2}\times\Gamma(\frac{1}{2})$$

$$=\frac{4}{\sqrt{\pi}}\times\frac{3}{2}\times\frac{1}{2}\times\sqrt{\pi}=3.$$

法二：用$\Gamma(\alpha)$函数的方法求解积分.

χ^2分布的性质

性质6.4.2 设$X \sim \chi^2(n_1)$,$Y \sim \chi^2(n_2)$,并且X与Y相互独立,则$X+Y \sim \chi^2(n_1+n_2)$.　　(χ^2分布的可加性)

设 $X_i \sim \chi^2(n_i)$, 并且 $X_i\ (i=1,2,\cdots,m)$ 相互独立,则$\sum_{i=1}^{m} X_i \sim \chi^2(n_1+n_2+\cdots+n_m)$.

性质6.4.3 设$\chi^2 \sim \chi^2(n)$,则对任意实数x有

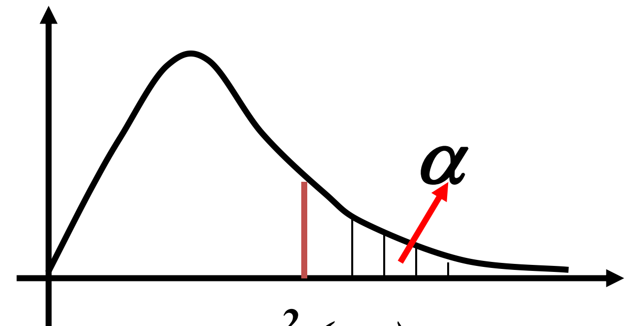
$$\lim_{n\to\infty} P\left\{\frac{\chi^2-n}{\sqrt{2n}} \le x\right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}}\,\mathrm{d}t.$$

这个性质说明当n很大时,$\frac{\chi^2-n}{\sqrt{2n}}$近似服从标准正态分布,也就是自由度为$n$的$\chi^2$分布近似于正态分布$N(n,2n)$.

χ^2分布的分位点（数）

对于给定的 $\alpha(0<\alpha<1)$，称满足条件：

$$P\{\chi^2 > \chi_\alpha^2(n)\} = \int_{\chi_\alpha^2(n)}^{+\infty} f(x)dx = \alpha$$



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

求$\chi_\alpha^2(n)$的值,可通过查表完成.

$$\boxed{\left\{\alpha, n, \chi_\alpha^2(n)\right\}\text{知二查一.}}$$

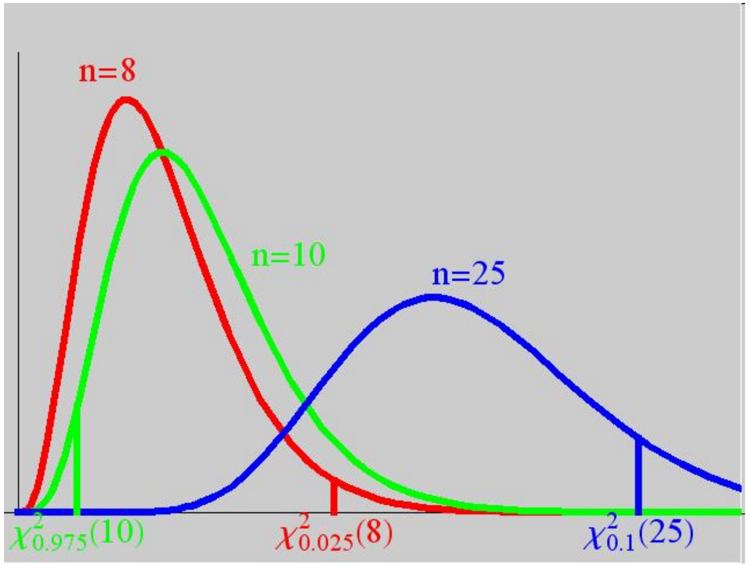
附表4，第$298-300$页-----χ^2分布表

例如：

$\chi^2_{0.025}(8) = 17.535,$ 附表4-3

$\chi^2_{0.975}(10) = 3.247,$ 附表4-1

$\chi^2_{0.1}(25) = 34.382.$ 附表4-4



附表4只详列到 n=45 为止.

服从卡方分布的统计量1

定理6.4.1 设$X \sim N(\mu,\sigma^2)$, $X_1, X_2, \cdots, X_n$是X的样本, 则随机变量

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

证:

因$X_1, X_2, \cdots, X_n$相互独立, 且都与总体X服从相同的正态分布$N(\mu, \sigma^2)$,

则$\dfrac{X_i-\mu}{\sigma} \sim N(0,1), i=1,2,\cdots,n,$

且$\dfrac{X_1-\mu}{\sigma}, \dfrac{X_2-\mu}{\sigma}, \cdots \dfrac{X_n-\mu}{\sigma}$相互独立.

由定义, 则

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

例6.4.1 设X_1，X_2，…，X_{16}为来自总体$N(0, 4)$的简单随机样本，求 $P\left\{\sum_{i=1}^{16} X_i^2 \le 77.476\right\}$.

解 因为$\dfrac{X}{2} \sim N(0,1)$, $\dfrac{X_i}{2} \sim N(0,1)$

所以 $\dfrac{1}{2^2}\sum_{i=1}^{16} X_i^2 \sim \chi^2(16)$.

则有

$$P\left\{\sum_{i=1}^{16} X_i^2 \le 77.476\right\} = P\left\{\frac{1}{2^2}\sum_{i=1}^{16} X_i^2 \le \frac{77.476}{2^2}\right\}$$

$$=1-P\left\{\frac{1}{2^2}\sum_{i=1}^{16} X_i^2 > \frac{77.476}{2^2}\right\} = 1-P\left\{\chi^2(16) > 19.369\right\}$$

$$=1-0.25 \ \boxed{=0.75}.$$

服从卡方分布的统计量2

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

定理6.4.2

设$X \sim N(\mu,\sigma^2)$, $X_1, X_2, \cdots, X_n$是X的样本, 样本均值和样本方差分别为$\overline{X}$和S^2, 则

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

(2) $\overline{X}$与S^2相互独立.

证明略.

χ^2分布的性质都要记住!

例6.4.2

设总体$X\sim N(\mu,3^2)$,其中μ为未知参数。从总体X中抽取容量为16的样本，求样本方差S^2小于16.5的概率。

解 **由定理6.4.2可知** $\chi^2=\dfrac{(16-1)S^2}{9}=\dfrac{5}{3}S^2\sim\chi^2(15).$

由此得所求概率

$$P\{S^2<16.5\}=P\{\frac{5}{3}S^2<\frac{5}{3}\times 16.5\}=P\{\chi^2(15)<27.5\}=1-P\{\chi^2(15)\geq 27.5\}$$

查附表4，当自由度n=15时，有 $\chi^2_{0.025}(15)=27.488,$

所以 $P\{\chi^2(15)\geq 27.5\}\approx 0.025,$

因此 $P\{S^2<16.5\}\approx 1-0.025=0.975.$

例6.4.3 设总体 $X \sim N(\mu,\sigma^2)$，从总体中抽取容量为n的样本，其样本方差为 S^2，求 $D(S^2)$．

解 由定理6.4.2可知 $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.

再由 χ^2分布的性质可得

$$D\left[\frac{(n-1)S^2}{\sigma^2}\right] = 2(n-1).$$

又

$$D\left[\frac{(n-1)S^2}{\sigma^2}\right] = \frac{(n-1)^2}{\sigma^4} D(S^2),$$

故

$$D(S^2) = \frac{2\sigma^4}{(n-1)}.$$

结论:

1. 对任意总体X

设$E(X)=\mu, D(X)=\sigma^2, X_1, X_2, \ldots, X_n$为样本,$\overline{X}$为样本均值，则

$$E(\overline{X})=\mu,$$

$$D(\overline{X})=\frac{\sigma^2}{n},$$

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

2. 若总体$X\sim N(\mu,\sigma^2)$,则还有

$$D(S^2)=\frac{2\sigma^4}{n-1}.$$

2. t-分布

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

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

服从自由度是n的 t 分布，

记作 $t \sim t(n)$.

表示一个随机变量

(1)自由度为n

(2)$T = \dfrac{X}{\sqrt{Y/n}}$

(3)$X \sim N(0,1), Y \sim \chi^2(n)$

t 分布由统计学家Gosset发现，又称学生氏(Student)分布.

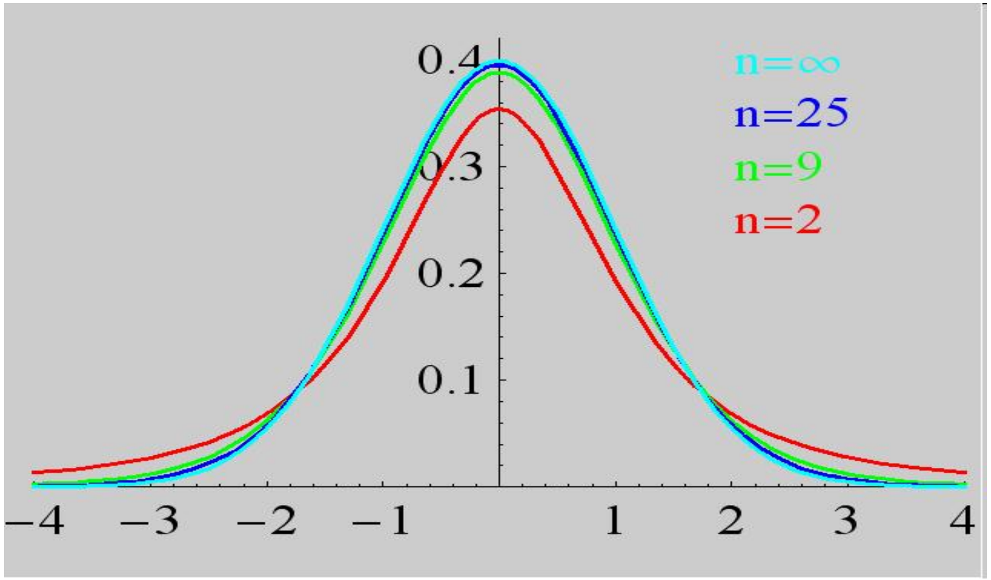
$t(n)$ 分布的概率密度函数为

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

t 分布的概率密度曲线如图

显然图形是关于 $t=0$ 对称的.

当 n 充分大时, 其图形类似于标准正态变量概率密度的图形.



因为$\lim\limits_{n\to\infty} h_n(t) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$,

所以当 n 足够大($n > 45$)时 t 分布近似于 $N(0,1)$ 分布,

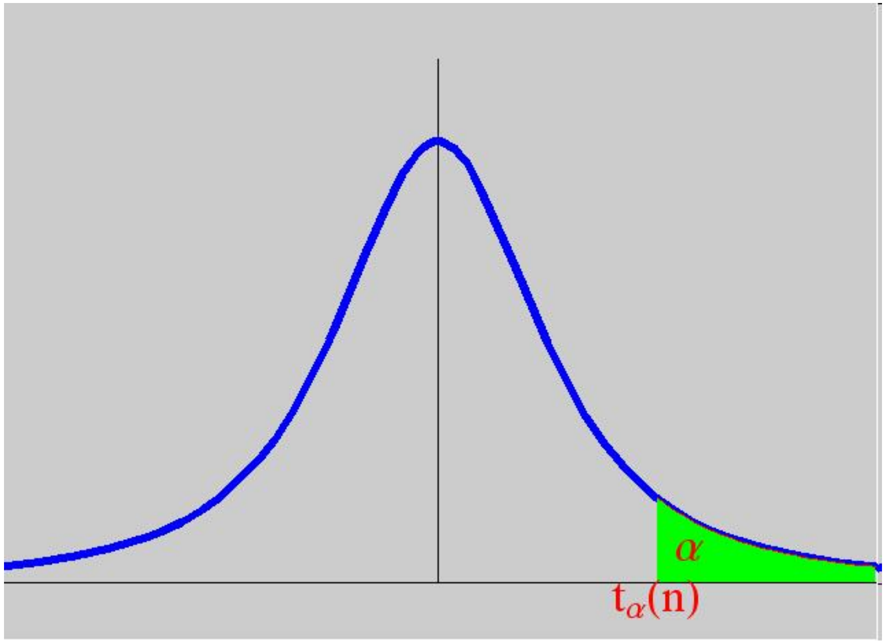
但对于较小的 n, t分布与$N(0,1)$分布相差很大.

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.



由分布的对称性知

$$t_{1-\alpha}(n)=-t_\alpha(n).$$

$$t_{0.5}(\mathrm{n})=0$$

当 $n>45$ 时, $t_\alpha(n)\approx u_\alpha$.

其中u_α为标准正态分布的上α分位点.

查附表3，第296-297页—— t 分布表

$\{\alpha, n, t_\alpha(n)\}$ 知二查一.

练习 查表求下列 $t_\alpha(n)$ 的值或概率的值。

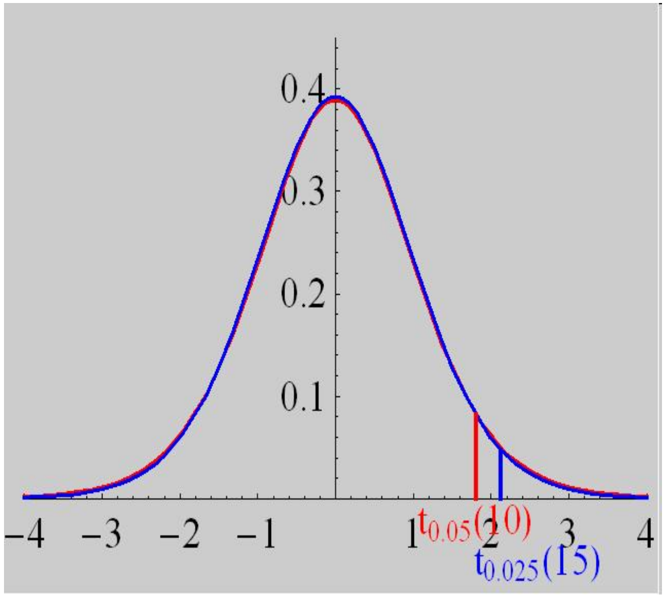
(1) $t_{0.05}(10) = \underline{1.8125}$;

(2) $t_{0.025}(15) = \underline{2.1315}$;

(3) $P\{t(8) > 1.3968\} = \underline{0.10}$,

（反查表）

(4) $t_{0.95}(10) = \underline{-1.8125}$.



服从t分布的统计量1

定理6.5.1 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

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

证明 因为 $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$, $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$,

且两者独立, 由 t 分布的定义知

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \Bigg/ \sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}} \sim t(n-1).$$

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

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

例6.5.1 设总体$X \sim N(\mu,\sigma^2)$,从总体X中抽取样本X_1, $X_2,\cdots$, X_n, X_{n+1}, 记$\overline{X}_n=\dfrac{1}{n}\sum_{i=1}^{n}X_i$, $S_n^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X}_n)^2$,

证明统计量$\sqrt{\dfrac{n}{n+1}}\cdot\dfrac{X_{n+1}-\overline{X}_n}{S_n}\sim t(n-1)$.

证明 因为$X_{n+1}\sim N(\mu,\sigma^2)$, $\overline{X}_n\sim N(\mu,\sigma^2/n)$,且$X_{n+1}$与$\overline{X}_n$相互独立,

所以$X_{n+1}-\overline{X}_n\sim N(0,\frac{n+1}{n}\sigma^2)$. 标准化可得, $u=\dfrac{X_{n+1}-\overline{X}_n}{\sqrt{\frac{n+1}{n}}\sigma}\sim N(0,1)$.

根据定理6.4.2, $\overline{X}_n$与S_n^2相互独立, $\chi^2=\dfrac{(n-1)S_n^2}{\sigma^2}\sim\chi^2(n-1)$.

由于X_{n+1}与S_n^2相互独立, 因此u与χ^2相互独立。

于是 $$\frac{u}{\sqrt{\chi^2/(n-1)}}=\frac{\dfrac{X_{n+1}-\overline{X}_n}{\sigma}\cdot\sqrt{\dfrac{n}{n+1}}}{\sqrt{\dfrac{(n-1)S_n^2}{\sigma^2(n-1)}}}=\frac{X_{n+1}-\overline{X}_n}{S_n}\cdot\sqrt{\frac{n}{n+1}}\sim t(n-1).$$

服从t分布的统计量2

定理6.5.2 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$\overline{X} = \dfrac{1}{n_1}\sum_{i=1}^{n_1} X_i$,

$\overline{Y} = \dfrac{1}{n_2}\sum_{i=1}^{n_2} Y_i$ 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

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

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

证明 根据定理6.3.2可得，$u=\dfrac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim N(0,1)$,

根据定理6.4.2可得

$$\chi_1^{\ 2}=\frac{(n_1-1)S_1^{\ 2}}{\sigma^2}\sim\chi^2(n_1-1),\quad \chi_2^{\ 2}=\frac{(n_2-1)S_2^{\ 2}}{\sigma^2}\sim\chi^2(n_2-1),$$

由两样本相互独立可得χ_1^2和χ_2^2相互独立，由χ^2分布的可加性得，

$$\chi_1^{\ 2}+\chi_2^{\ 2}=\frac{(n_1-1)S_1^{\ 2}+(n_2-1)S_2^{\ 2}}{\sigma^2}\sim\chi^2(n_1+n_2-2),$$

根据定理6.4.2可得$\overline{X}$与S_1^2相互独立，$\overline{Y}$与S_2^2相互独立，因此u与$\chi_1^2+\chi_2^2$相互独立，

于是$$\frac{u}{\sqrt{\frac{\chi_1^2+\chi_2^2}{n_1+n_2-2}}}=\frac{\dfrac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}}{\sqrt{\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{(n_1+n_2-2)\sigma^2}}}=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim t(n_1+n_2-2).$$

3. F分布

设 $X \sim \chi^2(n_1)$, $Y \sim \chi^2(n_2)$, 且X, Y 独立, 则称随机变量 $F = \dfrac{X/n_1}{Y/n_2}$服从自由度为 (n_1, n_2)的 F 分布, 记为 $F \sim F(n_1, n_2)$.

表示一个随机变量

(1)自由度为n_1, n_2

(2)$F = \dfrac{X/n_1}{Y/n_2}$

(3)$X \sim \chi^2(n_1), Y \sim \chi^2(n_2)$

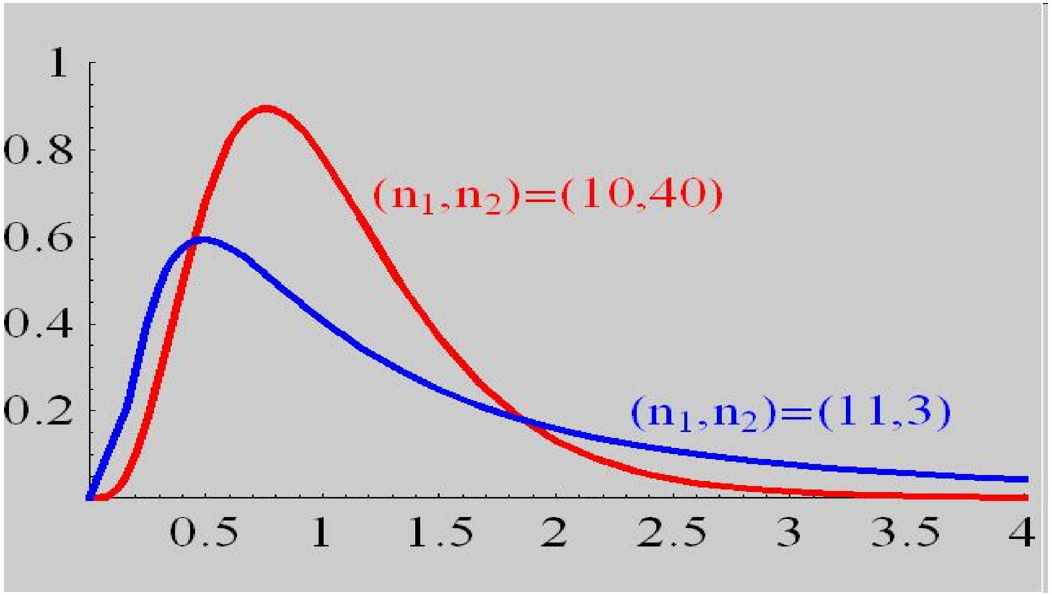
根据定义可知,

$$若F = \frac{X/n_1}{Y/n_2} \sim F(n_1, n_2), \quad 则\frac{1}{F} = \frac{Y/n_2}{X/n_1} \sim F(n_2, n_1).$$

$F(n_1, n_2)$分布的概率密度为

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

F分布的概率密度曲线如图



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.

查附表5，第301-310页—— F分布表

例6.5.2 设 $F(n_1, n_2)$分布的上α分位点满足

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

求 $F_\alpha(n_1, n_2)$的值, 可通过查表完成.

$F_{0.025}(8, 7) = \mathbf{4.90}$,

$F_{0.05}(30, 14) = \mathbf{2.31}$.

F分布的结论

(a)若 $T=\dfrac{X}{\sqrt{Y/n}}\sim t(n)$，则 $T^2=\dfrac{X^2}{Y/n}\sim F(1,n)$.

$(b)F_{1-\alpha}(m,n)=\dfrac{1}{F_\alpha(n,m)}$. 用于查表（当$\alpha$较大时）

证(b) 设$F\sim F(m,n)$，则

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

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

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

练习： $F_{0.95}(12,9)=?$

解： $F_{0.95}(12,9)=\dfrac{1}{F_{0.05}(9,12)}=\dfrac{1}{2.80}\approx 0.357.$

设$F\sim F(5,10)$，则$P\{F<\dfrac{1}{4.74}\}=$ 0.05 .

解： $P\left\{F<\dfrac{1}{4.74}\right\}=P\left\{\dfrac{1}{F}>4.74\right\}\overset{\text{设}}{=}\alpha \quad \left(\text{因为}\dfrac{1}{F}\sim F(10,5)\right)$

所以$4.74=F_{\alpha}(10,5)$ 反查表得到$\alpha=0.05.$

服从F分布的统计量1

定理6.6.1 设$X_1, X_2, \cdots, X_{n_1}$是来自正态总体$N(\mu_1, \boxed{\sigma_1^2})$的样本, $Y_1, Y_2, \cdots, Y_{n_2}$是来自正态总体$N(\mu_2, \boxed{\sigma_2^2})$的样本, 并且这两个样本相互独立, 则随机变量

$$F = \frac{n_2}{n_1} \cdot \frac{\sigma_2^2}{\sigma_1^2} \cdot \frac{\sum_{i=1}^{n_1} \left(X_i - \mu_1 \right)^2}{\sum_{j=1}^{n_2} \left(Y_j - \mu_2 \right)^2} \sim F\left(n_1, n_2 \right)$$

证:

根据定理6.4.1可得

$$\chi_1^2=\frac{1}{\sigma_1^2}\sum_{i=1}^{n_1}(X_i-\mu_1)^2\sim\chi^2(n_1),$$

$$\chi_2^2=\frac{1}{\sigma_2^2}\sum_{j=1}^{n_2}(Y_j-\mu_2)^2\sim\chi^2(n_2),$$

因X_i与Y_j($i=1,2,\cdots,n_1,\ j=1,2,\cdots,n_2$)相互独立，则$\chi_1^2$与$\chi_2^2$相互独立.

于是

$$F=\frac{\chi_1^2/n_1}{\chi_2^2/n_2}=\frac{n_2}{n_1}\cdot\frac{\sigma_2^2}{\sigma_1^2}\cdot\frac{\sum_{i=1}^{n_1}(X_i-\mu_1)^2}{\sum_{j=1}^{n_2}(Y_j-\mu_2)^2}\sim F(n_1,n_2).$$

服从F分布的统计量2

定理6.6.2 设从两个正态总体$N(\mu_1,\boxed{\sigma_1^2})$和$N(\mu_2,\boxed{\sigma_2^2})$中分别独立地各抽取一个样本,它们的样本容量分别为$n_1$和$n_2$,样本均值分别为$\overline{X}$和$\overline{Y}$,样本方差分别为$S_1^2$和$S_2^2$,则随机变量

$$F=\frac{\sigma_2^2}{\sigma_1^2}\cdot\frac{S_1^2}{S_2^2}\sim F(n_1-1,n_2-1).$$

证明

根据定理6.4.2可得

$$\chi_1^2=\frac{(n_1-1)S_1^2}{\sigma_1^2}\sim\chi^2(n_1-1),\quad \chi_2^2=\frac{(n_2-1)S_2^2}{\sigma_2^2}\sim\chi^2(n_2-1),$$

因为两个样本相互独立，所以S_1^2与S_2^2相互独立，从而χ_1^2与χ_2^2相互独立.

由定义可得

$$F=\frac{\chi_1^2/(n_1-1)}{\chi_2^2/(n_2-1)}=\frac{\sigma_2^2}{\sigma_1^2}\cdot\frac{S_1^2}{S_2^2}\sim F(n_1-1,n_2-1).$$

例6.5.3 分别独立地从正态总体 $X \sim N(\mu_1, 4)$ 和 $Y \sim N(\mu_2, 2)$ 中各抽取一个容量分别为10和9的样本，样本方差分依次为 S_1^2 和 S_2^2，求 $\dfrac{S_1^2}{S_2^2}$ 落在区间 $(0.366, 6.780)$ 内的概率。

解 根据定理6.6.2 $F = \dfrac{2S_1^2}{4S_2^2} = \dfrac{S_1^2}{2S_2^2} \sim F(9, 8)$,

则

$$P\left\{0.366 < \frac{S_1^2}{S_2^2} < 6.780\right\} = P\left\{0.183 < \frac{S_1^2}{2S_2^2} < 3.390\right\}$$

$$= P\{0.183 < F < 3.390\} = P\{F > 0.183\} - P\{F > 3.390\},$$

查表得 $F_{0.05}(9, 8) = 3.39$, 故 $P\{F > 3.39\} = 0.05$,

因0. 183离原点很近，设$P\{F > 0.183\} = 1 - \alpha$,

从而$0.183 = F_{1-\alpha}(9,8) = \dfrac{1}{F_\alpha(8,9)}$，故$F_\alpha(8, 9) = \dfrac{1}{0.183} = 5.47$,

查表得$F_{0.01}(8,9)=5.47$， 即$\alpha=0.01$,

故 $P\{F>0.183\}=1-0.01=0.99$,

又$P\{F>3.39\}=0.05$,

由此可得

$$P\left\{0.366<\frac{S_1^2}{S_2^2}<6.780\right\}=P\{F>0.183\}-P\{F>3.390\}$$

$$=0.99-0.05=0.94.$$

本章重点

1. 理解总体、样本、统计量的定义

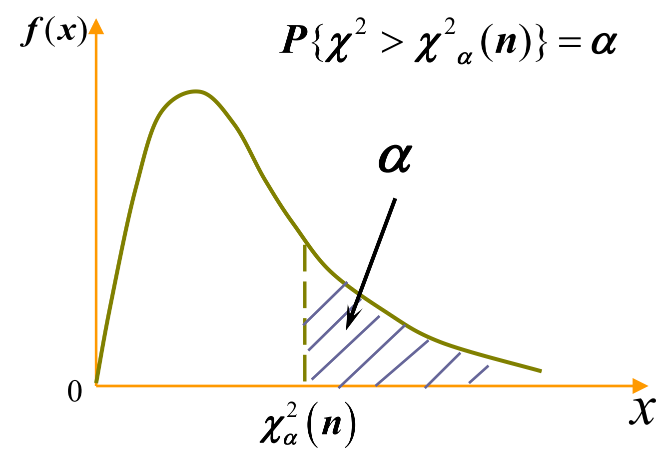
2. 掌握样本均值、样本方差、样本矩的定义、性质和计算

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2=\frac{1}{n-1}[\sum_{i=1}^{n}X_i^{\ 2}-n\overline{X}^2]$$

3. 熟练掌握 χ^2 分布、t分布和F分布的定义与性质、了解上α分位点的概念，会查表.

- $X_i\sim N(0,1),X_1,X_2,\cdots,X_n$相互独立，

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

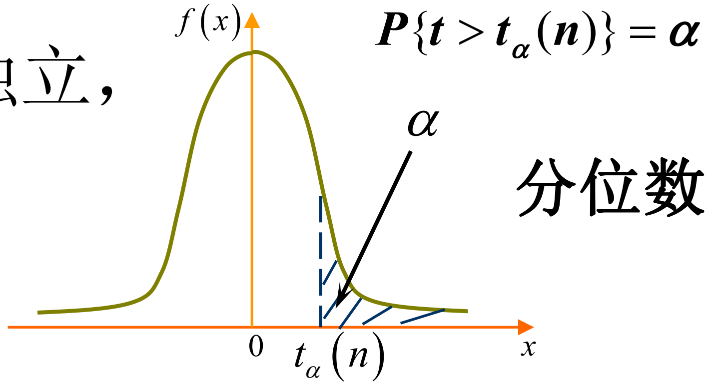


χ^2分布的分位数

若$X\ \sim\chi^2(n)$,则$E(X)=n,D(X)=2n$

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

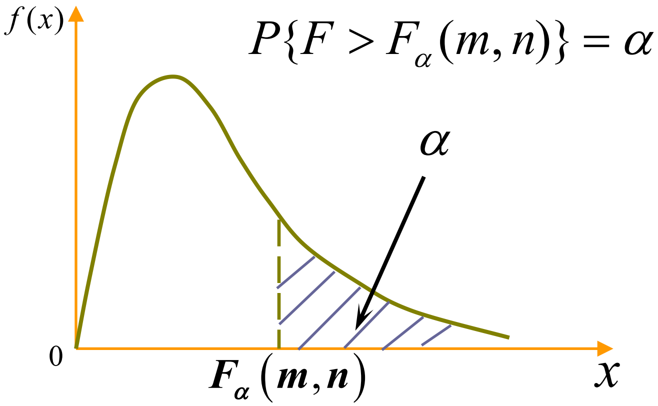
$$t = \frac{X}{\sqrt{Y/n}} \sim t(n)$$



t分布的分位数

- $X \sim \chi^2(m), Y \sim \chi^2(n)$, X与Y独立

$$F = \frac{X/m}{Y/n} \sim F(m,n)$$



F分布的分位数

4. 掌握正态总体的重要性质

设$X_1, X_2, \ldots, X_n$是总体$X \sim N(\mu, \sigma^2)$的一个样本，则

- $E(\bar{X}) = \mu, D(\bar{X}) = \dfrac{\sigma^2}{n}, E(S^2) = \sigma^2;$

  (此结论对一般总体也成立)

- $D(S^2) = \dfrac{2\sigma^4}{n-1};$

- $\bar{X} \sim N(\mu, \dfrac{\sigma^2}{n});$

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

- $\bar{X}$与S^2独立

双正态总体的重要性质

设$X_1, X_2, \ldots, X_{n_1}$是总体$X \sim N(\mu_1, \sigma_1^2)$的一个样本，

设$Y_1, Y_2, \ldots, Y_{n_2}$是总体$Y \sim N(\mu_2, \sigma_2^2)$的一个样本， 则

- $$\overline{X} - \overline{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right),$$

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

设$X_1, X_2, \ldots, X_n$是总体$X \sim N(\mu, \sigma^2)$的一个样本，则

服从卡方分布的统计量

- $$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 \sim \chi^2(n);$$
- $$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1);$$

设$X_1, X_2, \ldots, X_n$是总体$X \sim N(\mu, \sigma^2)$的一个样本，则

服从t分布的统计量

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

设$X_1, X_2, \ldots, X_{n_1}$是总体$X \sim N(\mu_1, \sigma_1^2)$的一个样本，

设$Y_1, Y_2, \ldots, Y_{n_2}$是总体$Y \sim N(\mu_2, \sigma_2^2)$的一个样本，则

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

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

设$X_1, X_2, \ldots, X_{n_1}$是总体$X \sim N(\mu_1, \sigma_1^2)$的一个样本，

设$Y_1, Y_2, \ldots, Y_{n_2}$是总体$Y \sim N(\mu_2, \sigma_2^2)$的一个样本，则

服从F分布的统计量

- $$F = \frac{n_2}{n_1} \cdot \frac{\sigma_2^2}{\sigma_1^2} \cdot \frac{\sum_{i=1}^{n_1} (X_i - \mu_1)^2}{\sum_{j=1}^{n_2} (Y_j - \mu_2)^2} \sim F(n_1, n_2)$$

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

哥塞特资料

William Sealey Gosset(1876-1937)

- 曾在牛津大学学习数学和化学
- 1899年开始在都柏林一家酿酒厂从事数据分析工作
- 1908年发现t分布
  以”student”的笔名发表了此成果
- t分布的发现打破了正态分布一统天下的局面
- 开创了小样本统计推断的新纪元

返回