§1.3 抽 样 分 布

在实际应用中，当我们从总体中抽取一个样本$(X_1, X_2, \cdots, X_n)$后，并不能直接应用它去对总体进行统计推断，这是因为样本虽然是从总体中获取的代表，含有总体性质的信息，但仍较分散。为了进行统计推断，需要把分散的信息集中起来，针对不同的研究目的，构造不同的样本函数，这种函数在统计学中称为统计量.

定义1 设$X_1, X_2, \cdots, X_n$是从总体X中抽取的容量为n的一个样本，如果由此样本构造一个函数$T(X_1, X_2, \cdots, X_n)$不依赖于任何未知参数，则称函数

$T(X_1, X_2, \cdots, X_n)$为一个统计量. 当获得样本的一组具体观测值$x_1, x_2, \cdots, x_n$后，称$T(x_1, x_2, \cdots, x_n)$为该统计统计量的一个观测值.

例1 设$X_1, X_2, \cdots, X_n$是从某总体X中抽取的一个样本，$E(X)$，$D(X)$未知，判断以下是否为统计量.

(1) $\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$,　　(2) $S_n^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$,

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

解：(1)(2)是统计量，　(3)(4)不是统计量.
因为(3)(4)依赖总体分布的未知参数.

一　常用的统计量

对于一维总体X，常用的统计量有

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

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

(3)样本标准差　$S=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2};$

(4)样本k阶矩　$A_k=\frac{1}{n}\sum_{i=1}^{n}x_i^k;$

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

此外，有顺序统计量、样本峰度和样本偏度等统计量. 对于二维总体(X,Y)，常用的统计量有

(6) 样本协方差

$$S_{XY}^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y});$$

(7) 样本相关系数

$$\rho_{XY}=\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\cdot\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}};$$

当样本取得观测值$x_1, x_2, \cdots, x_n$后，代入即可得到这些统计量的观测值.

寻找抽样分布一般有两种方法：

(1) 求出分布函数的精确表达式；

(2) 求其渐近分布.

只有在少数情况下，才能得到统计量的精确分布.

下面介绍数理统计中的三大分布.

二　数理统计的三大分布

(一) χ^2分布

设$X_1, X_2, \cdots, X_n$相互独立且都服从$N(0,1)$，则称随机变量

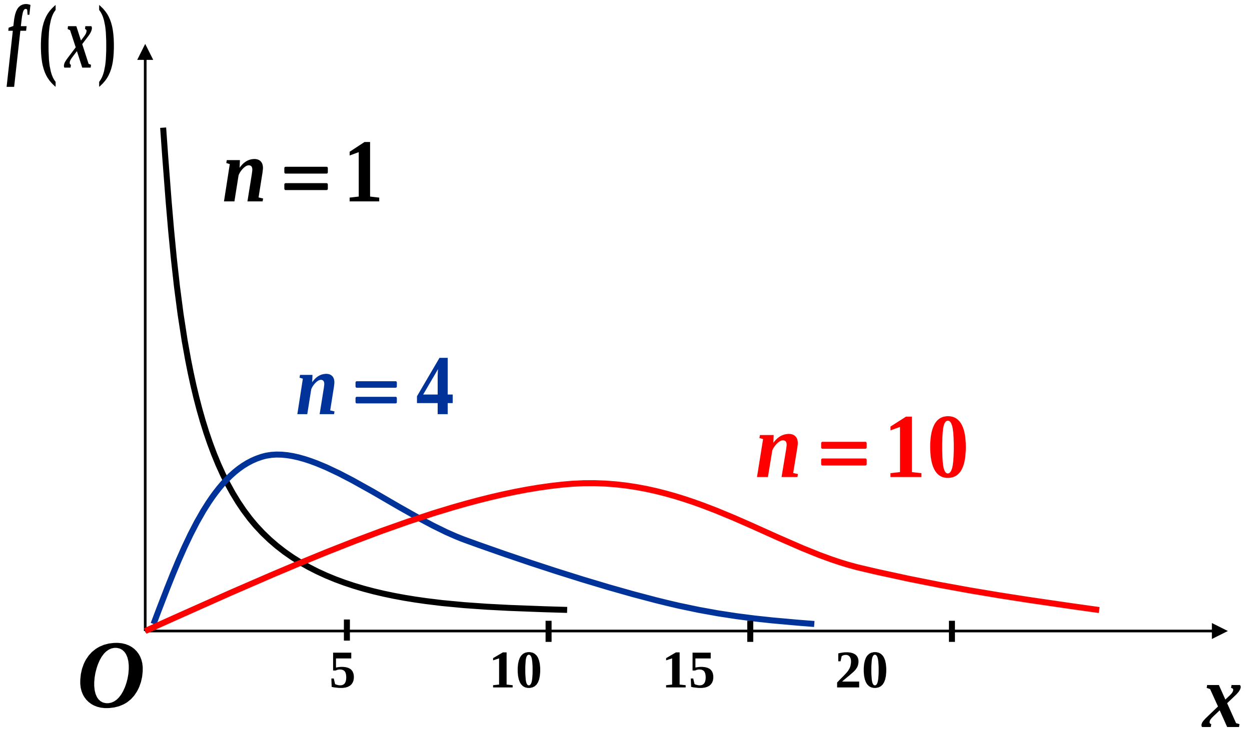
$$\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2$$

所服从的分布称为自由度为n的χ^2分布，记为$\chi^2(n)$.

其密度函数为

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

其中　$\Gamma(r) = \int_0^{\infty} x^{r-1} e^{-x} dx, \quad r > 0.$



χ^2分布的概率密度函数

χ^2分布的特点：

1. 分布的变量值始终为正；

2. 分布的形状取决于自由度n的大小，通常为不对称的正偏分布，但随着n的增大逐渐趋于对称；

χ^2分布的性质：

性质1 $E(\chi^2(n)) = n$，$D(\chi^2(n)) = 2n$;

证明： $X_i \sim N(0,1)$, $EX_i^2 = DX_i = 1$,

$D(X_i^2) = E(X_i^4) - [E(X_i^2)]^2 = 3 - 1 = 2$, 故

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

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

性质2　χ^2分布的可加性：

设$\chi_1^2 \sim \chi^2(n_1)$，$\chi_2^2 \sim \chi^2(n_2)$，并且相互独立，则

$$\chi_1^2 + \chi_2^2 \sim \chi^2(n_1 + n_2).$$

(此性质可以推广到多个随机变量的情形)

设$\chi_i^2 \sim \chi^2(n_i)$，并且$\chi_i^2 (i = 1, \cdots, m)$相互独立, 则

$$\sum_{i=1}^{m} \chi_i^2 \sim \chi^2(n_1 + n_2 + \cdots + n_m).$$

证明：利用卷积公式

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z - x) dx$$

即可证明. 亦可利用特征函数证明.

性质3 设$\chi^2 \sim \chi^2(n)$，则对任意x，有

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

证明 由假设$\chi^2 = \sum_{i=1}^{n} X_i^2$, 其中$X_1, X_2, \cdots, X_n$独立且每个$X_i \sim N(0,1)$，因而$X_1^2, X_2^2, \cdots, X_n^2$独立同分布，且

$$E(X_i^2) = 1, \qquad D(X_i^2) = 2. \qquad (i = 1, 2, \cdots, n)$$

由中心极限定理得

$$\lim_{n\to\infty} P\{\frac{\chi^2 - n}{\sqrt{2n}} \le x\} = \lim_{n\to\infty} P\{\frac{\sum_{i=1}^{n} X_i^2 - n\mu}{\sqrt{n}\sigma} \le x\}$$

$$= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

即χ^2分布的极限分布是正态分布，当n很大时

$$\chi^2(n) \overset{\text{近似}}{\sim} N(n, 2n).$$

(二) t分布

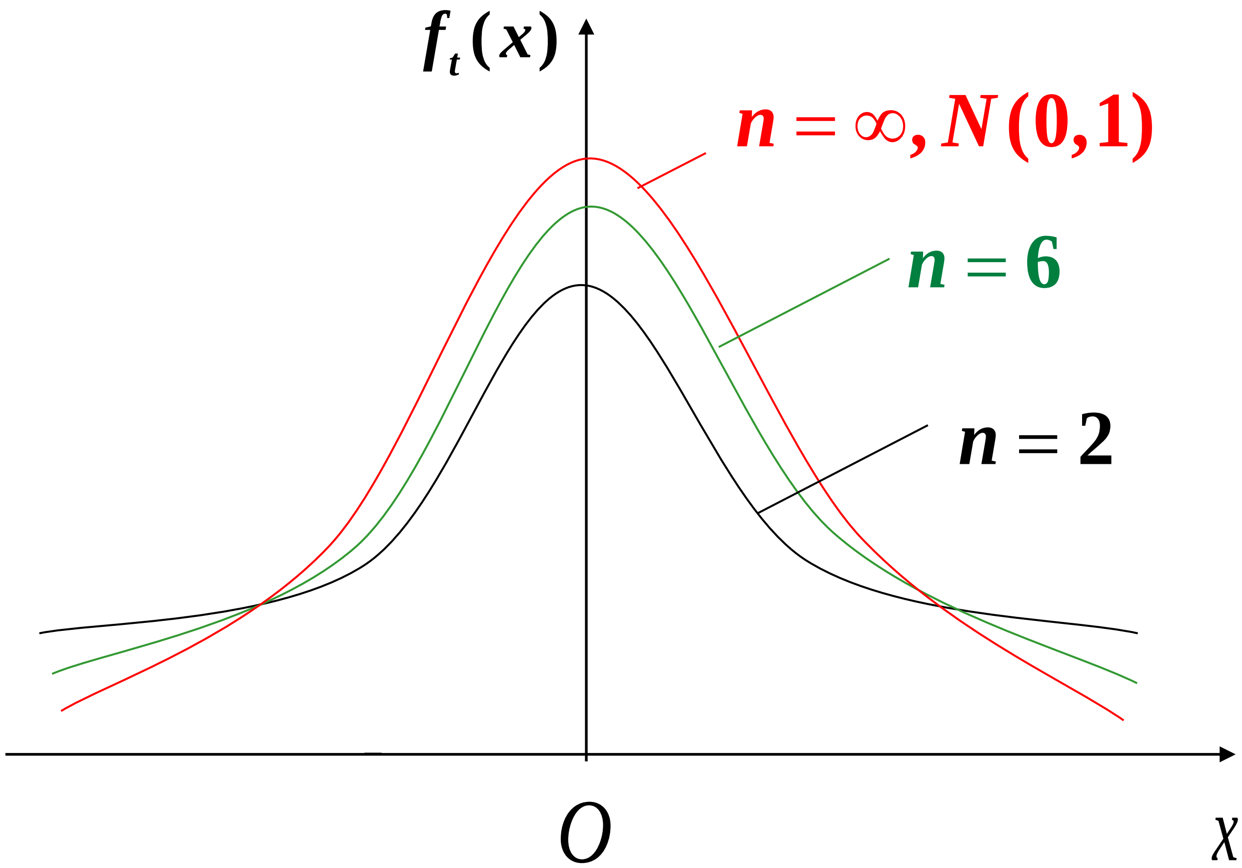
设随机变量$X \sim N(0,1)$, $Y \sim \chi^2(n)$，且X与Y独立，则随机变量

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

所服从的分布称为自由度为n的t分布，记为$t(n)$.

其密度函数为

$$f_t(x;n) = \frac{\Gamma[(n+1)/2]}{\Gamma(n/2)\sqrt{n\pi}}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < x < +\infty.$$



t分布的密度函数： 低峰、厚尾

t分布的性质：

1. 密度函数$f(x,n)$是偶函数，且

$$\lim_{n\to\infty} f(x,n)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}=\varphi(x).$$

此说明：t分布的极限分布是标准正态分布.

2. $t(1)$是标准柯西分布，它的均值不存在,

当$n\geq 2$时，t分布的数学期望 $E(T)=0$,

当$n\geq 3$时，t分布的方差 $D(T)=\dfrac{n}{n-2}$.

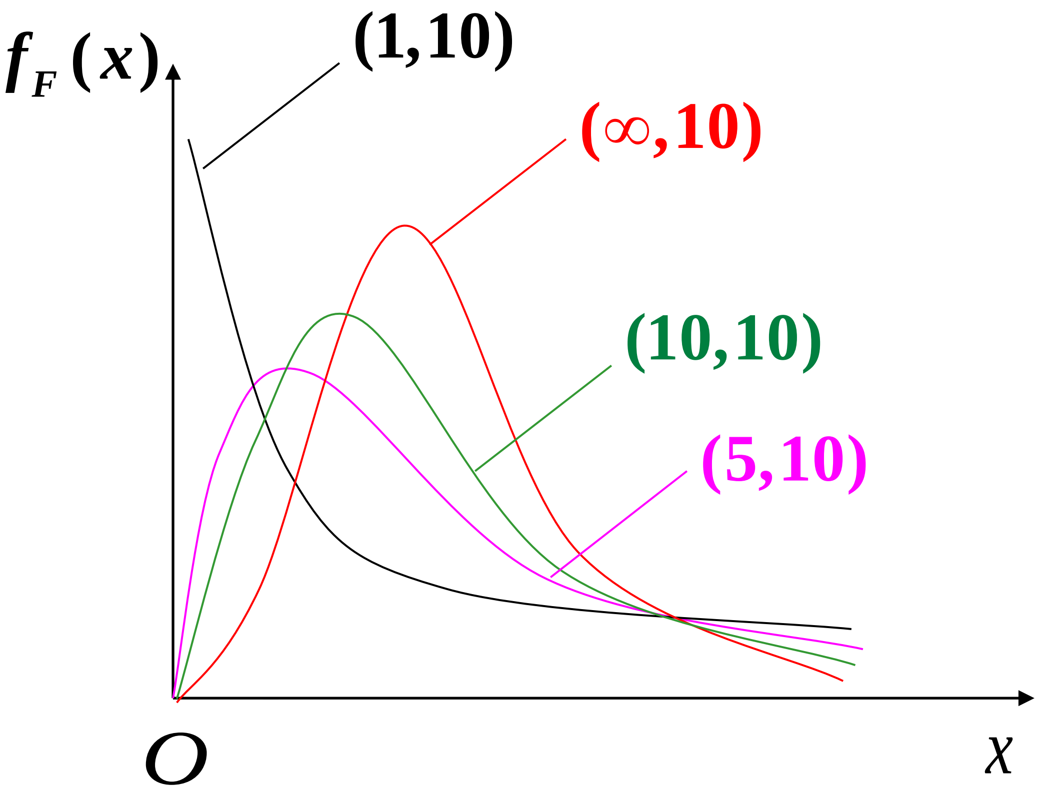
(三) F分布

设随机变量$X \sim \chi^2(m), Y \sim \chi^2(n)$，且$X$与$Y$独立，则随机变量

$$F = \frac{X/m}{Y/n}$$

的分布称为自由度为m与n的F分布, 记为$F \sim F(m, n)$, 其中m为分子自由度，n为分母自由度. 其密度函数为

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



F分布概率密度函数

首先，求出$Z=X/Y$的密度函数，再求$F=nZ/m$的密度函数.

$$p_Z(z)=\int_0^{+\infty} y p_1(zy)p_2(y)dy$$

$$=\frac{z^{\frac{m}{2}-1}}{\Gamma(m/2)\Gamma(n/2)2^{\frac{m+n}{2}}}\int_0^{+\infty} y^{\frac{m+n}{2}-1}e^{-\frac{y}{2}(1+z)}dy$$

做变换令$u=y(1+z)/2$，　于是

$$p_Z(z)=\frac{z^{\frac{m}{2}-1}(1+z)^{-\frac{m+n}{2}}}{\Gamma(m/2)\Gamma(n/2)}\int_0^{+\infty} u^{\frac{m+n}{2}-1}e^{-u}du$$

$$=\frac{\Gamma\left[(m+n)/2\right]}{\Gamma(m/2)\Gamma(n/2)}z^{\frac{m}{2}-1}(1+z)^{-\frac{m+n}{2}},z>0.$$

求$F=nZ/m$的密度函数.

$$f_F(x)=f_Z(\frac{m}{n}x)\cdot\frac{m}{n}$$

$$=\frac{\Gamma\left[(m+n)/2\right]}{\Gamma(m/2)\Gamma(n/2)}(\frac{m}{n}x)^{\frac{m}{2}-1}(1+\frac{m}{n}x)^{-\frac{m+n}{2}}\cdot\frac{m}{n}$$

$$=\frac{\Gamma\left[(m+n)/2\right]}{\Gamma(m/2)\Gamma(n/2)}m^{\frac{m}{2}}n^{\frac{n}{2}}\frac{x^{\frac{m}{2}-1}}{(mx+n)^{\frac{m+n}{2}}}$$

这就是自由度为m与n的F分布的密度函数.

F分布的性质：

性质1　若$X \sim F(m,n)$，则$1/X \sim F(n,m)$;

性质2　若$X \sim t(n)$，则 $X^2 \sim F(1,n)$;

性质3　$E(F)=\dfrac{n}{n-2} \quad (n>2)$,

$$D(F)=\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}, \quad (n>4).$$

$F-$分布是为纪念英国著名统计学家费歇（R.A. Fisher,1890－1962）而命名的．它是数理统计的重要分布之一.

【例1】设随机变量X与Y相互独立，$X \sim N(0,16)$, $Y \sim N(0,9)$, $X_1, X_2, \cdots, X_9$与$Y_1, Y_2, \cdots, Y_{16}$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　因为$X_1 + X_2 + \cdots + X_9 \sim N(0, 9 \times 16)$,

所以　$$\frac{1}{12}(X_1 + X_2 + \cdots + X_9) \sim N(0, 1).$$

由于　$\frac{Y_i}{3} \sim N(0,1)$, $\frac{1}{9}\sum_{i=1}^{16} Y_i^2 \sim \chi^2(16)$,

从而 $\dfrac{X_1+X_2+\cdots+X_9}{\sqrt{Y_1^2+Y_2^2+\cdots+Y_{16}^2}}=\dfrac{(X_1+X_2+\cdots+X_9)/12}{\sqrt{\frac{1}{9}\sum\limits_{i=1}^{16}Y_i^2\Big/16}}\sim t(16)$.

例2 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$\chi^2$分布.

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

$$(X_1+X_2+X_3)\big/\sqrt{3}\,,\,(X_4+X_5+X_6)\big/\sqrt{3}\sim N(0,1).$$

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

$=\dfrac{1}{3}Y\sim\chi^2(2)$. 因此 $c=1/3$.

三　分位数(点)

定义　设X的分布函数为$F(x)$，　满足等式

$$P(X > x_\alpha) = 1 - F(x_\alpha) = \alpha$$

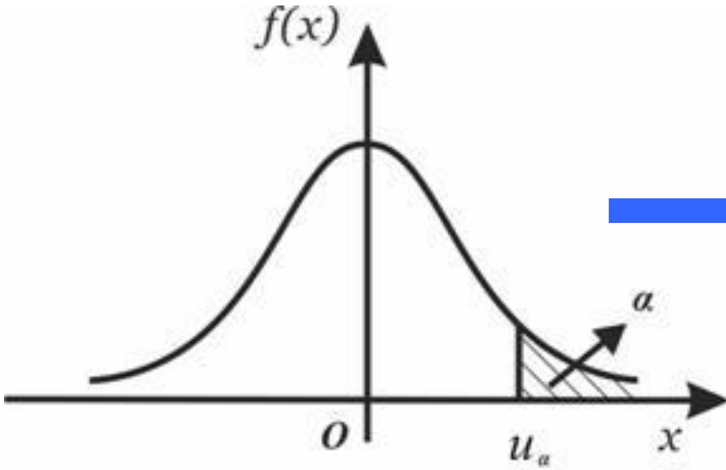
的实数x_α称为随机变量X的上α分位数.　$0 < \alpha < 1$.

注：若$F(x)$不是严格递增的连续函数时，为保证x_α的存在性和唯一性，定义改为

$$x_\alpha = \inf\left\{x : 1 - F(x) \le \alpha\right\}.$$

标准正态分布，$\chi^2(n)$，$t(n)$，$F(m, n)$的上α分位数分别记为u_α，$\chi^2_\alpha(n)$，$t_\alpha(n)$，$F_\alpha(m, n)$

如下图所示：



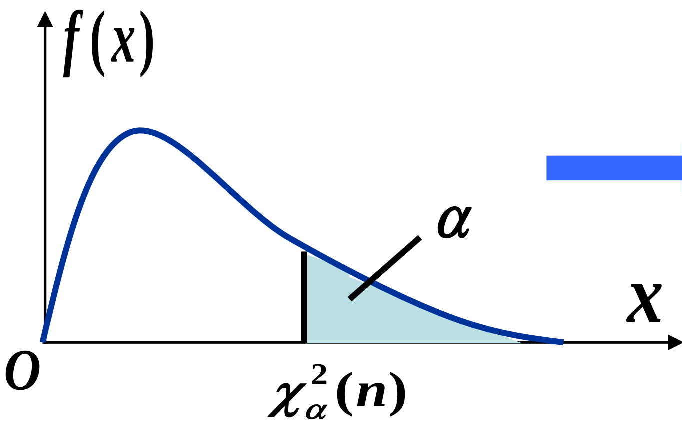
$u_{0.05}=1.645$

$u_{0.025}=1.96$

$u_{0.005}=2.575$

常用数字

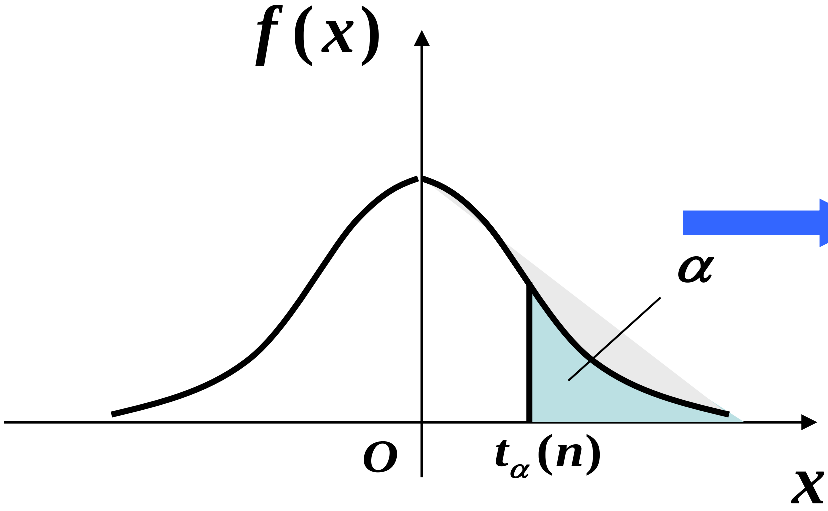
性质：$u_{1-\alpha}=-u_\alpha$



当$n<45$时, 对某些特殊的α, 可查表得到$\chi^2_\alpha(n)$.

性质：当$n\geq45$时，$\chi^2_\alpha(n)\approx\frac{1}{2}(u_\alpha+\sqrt{2n-1})^2$.

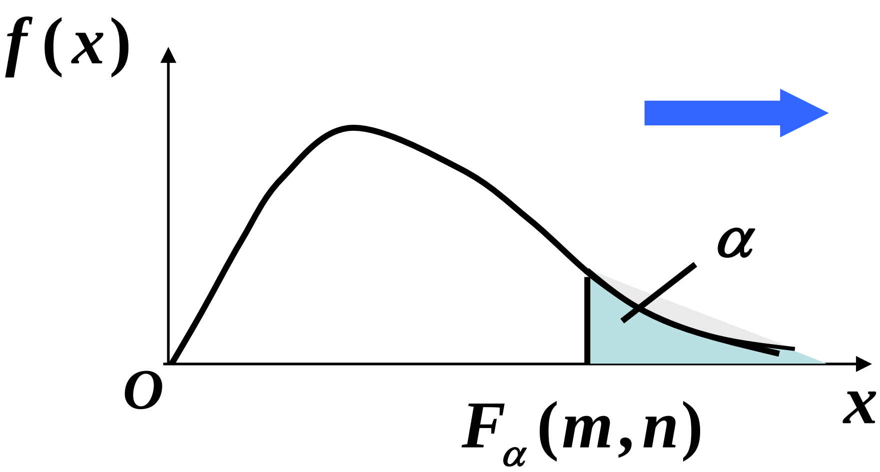
费歇证明：$\sqrt{2\chi^2(n)}\xrightarrow[n\geq45]{\text{近似}}N(\sqrt{2n-1},1)$.



$n < 45$时，查表得$t_\alpha(n)$.

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

性质：$t_{1-\alpha} = -t_\alpha$



对某些n,α．查表得$F_\alpha(m,n)$.

性质：$F_\alpha(m,n) = \dfrac{1}{F_{1-\alpha}(n,m)}$.

用来求解α较大时的分位数.

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

设$F\sim F(m,n)$，则 $1/F\sim F(n,m)$

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

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

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

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

四　正态总体的抽样分布定理

正态总体是最常见的总体，本节介绍的几个抽样分布定理均对正态总体而言. 这里我们主要掌握定理的结论，对定理的证明不作要求.

定理1　设$X_1, X_2, \cdots, X_n$是来自正态总体$X \sim N(\mu,\sigma^2)$的一组样本，则有下列结论：

$$(1)\ \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \frac{\sigma^2}{n}) \text{或} \frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim N(0,1);$$

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

$$(3)\ \bar{X} \text{与} S^2 \text{独立，且} \frac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t(n-1).$$

证明： 记 $X=(X_1,\cdots,X_n)^T$，则有

$$E(X)=(\mu,\cdots,\mu)^T,\ D(X)=\sigma^2 I$$

取一n阶正交矩阵A，其第1行的每个元素均为$1/\sqrt{n}$. 如

$$A=\begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{2\cdot 1}} & -\frac{1}{\sqrt{2\cdot 1}} & 0 & \cdots & 0 \\ \frac{1}{\sqrt{3\cdot 2}} & \frac{1}{\sqrt{3\cdot 2}} & -\frac{2}{\sqrt{3\cdot 2}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & -\frac{n-1}{\sqrt{n(n-1)}} \end{pmatrix}$$

令$Y=AX$，则由多维正态分布的性质知Y仍服从n维正态分布，其均值和方差分别为

$$EY=A\cdot EX=(\sqrt{n}\mu,0,\cdots,0)^T,$$

$$D(Y)=A\cdot D(X)\cdot A^T=A\cdot\sigma^2 I\cdot A^T=\sigma^2 AA^T=\sigma^2 I.$$

由此$Y=(Y_1,\cdots,Y_n)^T$的各分量相互独立，且都服从正态分布

$$Y_1\sim N(\sqrt{n}\mu,\sigma^2),\quad Y_2,\cdots,Y_n\sim N(0,\sigma^2)$$

由于$\bar{X}=Y_1/\sqrt{n}$，故$\bar{X}$与$Y_2,\cdots,Y_n$相互独立, 且

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

标准化得

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

由于

$$\sum_{i=1}^{n} Y_i^2 = Y^T Y = X^T A^T A X = \sum_{i=1}^{n} X_i^2$$

所以

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

$$= \sum_{i=1}^{n} Y_i^2 - Y_1^2 = \sum_{i=2}^{n} Y_i^2 .$$

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

由于$\bar{X}$与$Y_2, \cdots, Y_n$相互独立. 且

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

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

因而$\bar{X}$与S^2相互独立，再根据t分布的定义得

$$\frac{\sqrt{n}(\bar{X}-\mu)}{S} = \frac{\dfrac{\sqrt{n}(\bar{X}-\mu)}{\sigma}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2} \Big/ n-1}} \sim t(n-1).$$

定理2 设$X_1, X_2, \cdots, X_m$是来自正态总体$X \sim N(\mu_1, \sigma_1^2)$的一组样本, $Y_1, Y_2, \cdots, Y_n$是来自正态总体$Y \sim N(\mu_2, \sigma_2^2)$的一组样本，且两组样本相互独立. 记$\bar{X}, \bar{Y}$分别是两组的样本均值，$S_X^2$与$S_Y^2$分别是样本方差，则

(1) $$F = \frac{S_X^2 / S_Y^2}{\sigma_1^2 / \sigma_2^2} \sim F(m-1, n-1),$$

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

(3) 当$\sigma_1^2 = \sigma_2^2 = \sigma^2$时,

$$\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_W\sqrt{1/m+1/n}} \sim t(m+n-2).$$

其中 $S_W^2=\dfrac{(m-1)S_X^2+(n-1)S_Y^2}{m+n-2}$.

(4) 当$m=n$时

$$\frac{\sqrt{n}[(\bar{X}-\bar{Y})-(\mu_1-\mu_2)]}{S} \sim t(n-1).$$

其中

$$S^2=S_X^2+S_Y^2-2S_{XY}^2.$$

证明: (1) 由两样本独立可知

S_X^2与S_Y^2相互独立，且

$$\frac{(m-1)S_X^2}{\sigma_1^2} \sim \chi^2(m-1), \quad \frac{(n-1)S_Y^2}{\sigma_2^2} \sim \chi^2(n-1).$$

由F的定义知

$$F = \frac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2} \sim F(m-1, n-1).$$

(2) 由$\bar{X} \sim N(\mu_1, \frac{\sigma_1^2}{m})$, $\bar{Y} \sim N(\mu_2, \frac{\sigma_2^2}{n})$, $\bar{X}$与$\bar{Y}$独立

$$\Rightarrow \bar{X} - \bar{Y} = N(\mu_1 - \mu_2,\ \sigma_1^2/m + \sigma_2^2/n)$$

$$\Rightarrow \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}} \sim N(0,1).$$

(3) 由定理1知:

$$\frac{(m-1)S_X^2}{\sigma^2} \sim \chi^2(m-1), \quad \frac{(n-1)S_Y^2}{\sigma^2} \sim \chi^2(n-1), \text{独立}$$

$$\xrightarrow{\chi^2\text{分布可加性}}$$

$$\frac{(m+n-2)S_w^2}{\sigma^2} = \frac{(m-1)S_X^2 + (n-1)S_Y^2}{\sigma^2} \sim \chi^2(m+n-2)$$

又$\bar{X}-\bar{Y}$与S_w^2相互独立，根据t分布的定义即得所证.

(4) 当$m=n$时 令

$$Z_i = X_i - Y_i, \quad i=1,2,\cdots,n.$$

则 $Z_1, Z_2, \cdots, Z_n$独立同分布,

$$Z_i \sim N(\mu_1-\mu_2, \sigma_1^2+\sigma_2^2),$$

根据定理1得

$$\frac{\sqrt{n}\left[\bar{Z}-(\mu_1-\mu_2)\right]}{S_Z}\sim t(n-1).$$

由于 $\bar{Z}=\bar{X}-\bar{Y},$

$$S_Z^2=\frac{1}{n-1}\sum_{i=1}^{n}(Z_i-\bar{Z})^2$$

$$=\frac{1}{n-1}\sum_{i=1}^{n}\left[(X_i-Y_i)-(\bar{X}-\bar{Y})\right]^2$$

$$=S_X^2+S_Y^2-2S_{XY}^2=S^2.$$

所以

$$\frac{\sqrt{n}[(\bar{X}-\bar{Y})-(\mu_1-\mu_2)]}{S}\sim t(n-1).$$

定理3 (柯赫伦定理) 设$X_1, X_2, \cdots, X_n$独立，都服从$N(0,1)$.

$$Q = Q_1 + Q_2 + \cdots + Q_k = \sum_{i=1}^{n} X_i^2$$

其中Q_i是秩为n_i的$X_1, X_2, \cdots, X_n$的二次型. 则Q_i相互独立, 且$Q_i \sim \chi^2(n_i)(i = 1, \cdots, k) \Leftrightarrow \sum_{i=1}^{k} n_i = n$.

证明： 必要性

由$Q_1, \cdots, Q_k$相互独立，且$Q_i \sim \chi^2(n_i)$

$$\Rightarrow Q = Q_1 + Q_2 + \cdots + Q_k \sim \chi^2(\sum_{i=1}^{k} n_i) = \chi^2(n).$$

$$\Rightarrow \sum_{i=1}^{k} n_i = n.$$

充分性 根据线性代数知识，二次型$Q_1,\cdots,Q_k$可化为下列标准形式

$$\begin{cases} Q_1 = b_{11}Y_{11}^2 + b_{12}Y_{12}^2 + \cdots + b_{1n_1}Y_{1n_1}^2 \\ Q_2 = b_{21}Y_{21}^2 + b_{22}Y_{22}^2 + \cdots + b_{2n_2}Y_{2n_2}^2 \\ \qquad\cdots\cdots\cdots\cdots\cdots\cdots \\ Q_k = b_{k1}Y_{k1}^2 + b_{k2}Y_{k2}^2 + \cdots + b_{11}Y_{kn_k}^2 \end{cases}$$

其中Y_{ij}都是$X_1,X_2,\cdots,X_n$的线性组合，$b_{ij}=1$或-1.

记$Y=(Y_{11},\cdots,Y_{1n_1},Y_{21}\cdots,Y_{2n_2},\cdots,Y_{k1},\cdots,Y_{kn_k})^T$

—n维列向量,

$B=diag\{b_{11},\cdots,b_{1n_1},b_{21}\cdots,b_{2n_2},\cdots,b_{k1},\cdots,b_{kn_k}\}$

—n阶对角矩阵.

记$X=(X_1,X_2,\cdots,X_n)^T$，$Y=AX$．则

$$X^TX=\sum_{i=1}^{n}X_i^2=Q=\sum_{i=1}^{k}Q_i=\sum_{i=1}^{k}\sum_{j=1}^{n_i}b_{ij}Y_{ij}^2=Y^TBY$$

$$=X^TA^TBAX$$

$$\Rightarrow A^TBA=I_n\Rightarrow B=(AA^T)^{-1}\text{是正定矩阵}\Rightarrow b_{ij}=1$$

$$\Rightarrow B=I_n\Rightarrow AA^T=B^{-1}=I_n\Rightarrow A\text{是正交矩阵.}$$

因为$X\sim N_n(0,I_n)$,所以

$$Y=AX\sim N_n(0,AA^T)=N_n(0,I_n).$$

因而Y的n个分量独立同分布于$N(0,1)$．根据Q_i的表达式知

$$Q_1,\cdots,Q_k\text{相互独立，且 }Q_i\sim\chi^2(n_i)(i=1,\cdots,k).$$

例3：设总体X服从正态分布$N(\mu,\sigma^2)$，从总体中抽取样本$X_1,X_2,\cdots X_{n+1}$，记

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

证明：
$$\sqrt{\frac{n}{n+1}}\frac{X_{n+1}-\bar{X}_n}{S_n}\sim t(n-1)$$

证明　$\bar{X}_n$与X_{n+1}相互独立，且

$$\bar{X}_n\sim N(\mu,\sigma^2/n)\ ,\quad X_{n+1}\sim N(\mu,\sigma^2)$$

故
$$X_{n+1}-\bar{X}_n\sim N(0,(1+1/n)\sigma^2),$$

$$\sqrt{\frac{n}{n+1}}\frac{X_{n+1}-\bar{X}_n}{\sigma}\sim N(0,1).$$

再根据$X_{n+1}-\bar{X}_n$与S_n独立，及t分布的定义即得证.

练习题:

调节一个装瓶机使其对每个瓶子的灌装量均值为μ盎司，通过观察这台装瓶机对每个瓶子的灌装量服从标准差$\sigma=1.0$盎司的正态分布. 随机抽取由这台机器灌装的9个瓶子形成一个样本，并测定每个瓶子的灌装量. 试确定样本均值偏离总体均值不超过0.3盎司的概率有多大?

练习题答案：

解：由题意知，$\bar{X} \sim N(\mu, \dfrac{\sigma^2}{n})$，故$\bar{X} \sim N(\mu, \dfrac{1}{9})$

$$P(\left|\bar{X} - \mu\right| \le 0.3) = P(\frac{-0.3}{1/3} \le \frac{\bar{X} - \mu}{1/3} \le \frac{0.3}{1/3})$$

$$= \Phi(0.9) - \Phi(-0.9)$$

$$= 2\Phi(0.9) - 1$$

$$= 2 \times 0.8159 - 1 = 0.6318.$$

在练习题中，如果希望$\bar{X}$与μ的偏差在0.3盎司之内的概率达到0.95，应当抽取多大的样本？

解： $\bar{X} \sim N(\mu, \frac{1}{n})$，$P(|\bar{X}-\mu| \le 0.3) = 0.95$

即 $$P(\frac{|\bar{X}-\mu|}{1/\sqrt{n}} \le \frac{0.3}{1/\sqrt{n}}) \le 0.95.$$

而 $P(\frac{|\bar{X}-\mu|}{1/\sqrt{n}} \le \frac{0.3}{1/\sqrt{n}}) = 2\Phi(0.3\sqrt{n}) - 1.$ 所以

$$2\Phi(0.3\sqrt{n}) - 1 \ge 0.95, \quad \Phi(0.3\sqrt{n}) \ge 0.975$$

查表得： $0.3\sqrt{n} \ge 1.96,$ 即$n \ge 42.68.$

所以应当抽取容量至少为43的样本.

例4 设$X_1, X_2, \cdots, X_6$是来自正态总体$N(2,3)$的样本. 求b使$P\{\sum_{i=1}^{6}(X_i-2)^2 \le b\}=0.95.$

解: $\dfrac{X_i-2}{\sqrt{3}} \sim N(0,1) \quad (i=1,\cdots,6)$且相互独立,

$$\Rightarrow \sum_{i=1}^{6}\left(\frac{X_i-2}{\sqrt{3}}\right)^2 \sim \chi^2(6)$$

$$0.95=P\{\sum_{i=1}^{6}(X_i-2)^2 \le b\}=P\{\sum_{i=1}^{6}\left(\frac{X_i-2}{\sqrt{3}}\right)^2 \le \frac{b}{3}\}$$

$$=1-P\{\chi^2(6) \ge \frac{b}{3}\}$$

即$P\{\chi^2(6) > b/3\}=0.05$. 查表知,

$P\{\chi^2(6) > 12.592\}=0.05, \quad b/3=12.592 \Rightarrow b=37.776$

例5 设两正态总体X,Y的方差分别为$\sigma_1^2=12,\sigma_2^2=18,$ 在X,Y中分别取出样本容量为$n_1=61,n_2=31$的样本, 两样本独立, 样本方差为S_1^2,S_2^2,求$P\{S_1^2/S_2^2>1.16\}.$

解： $$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}=\frac{S_1^2/12}{S_2^2/18}\sim F(61-1,31-1)=F(60,30)$$

$$P\{S_1^2/S_2^2>1.16\}=P\{\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}>\frac{1.16}{12/18}=1.74\}$$

查表知, $F_{0.05}(60,30)=1.74$

$$\Rightarrow\quad P\{S_1^2/S_2^2>1.16\}=0.05$$

本章总结:

1) **数理统计基本概念:**

- 总体与个体
- 样本 (特征: 独立同分布)
- 统计量 (特征: 无未知参数)
- 常见统计量:
  - 样本均值
  - 样本方差
  - ……

2) **抽样分布:**

- $\chi^2(n)$ 分布
- $t(n)$ 分布
- $F(m,n)$ 分布

$\chi^2(n)$、$t(n)$、$F(m,n)$ 分布: 定义、性质、分位点

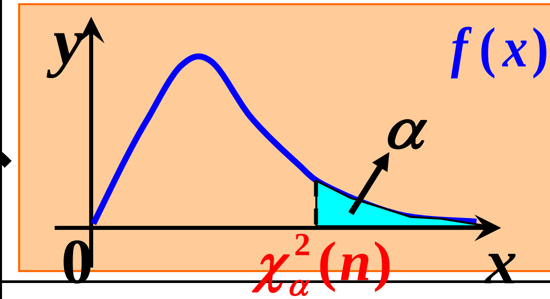
- 正态总体的抽样分布
  - $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$,
  - $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

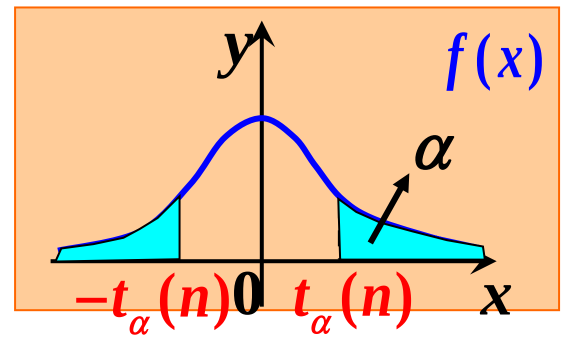
常见统计量及正态总体抽样分布

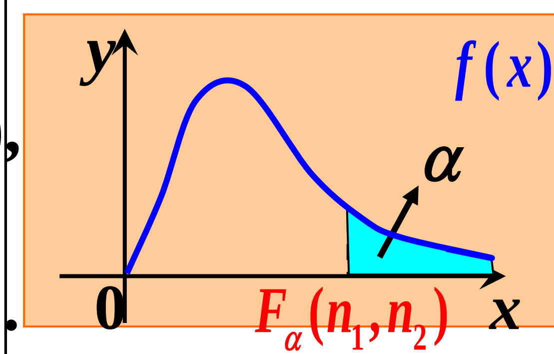
名称	定义	性质	正态抽样分布
样本均值	$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$	$E(\overline{X})=E(X)$,	$\overline{X}\sim N(\mu,\frac{\sigma^2}{n})$;
样本方差	$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ $=\frac{1}{n-1}(\sum_{i=1}^{n}X_i^2-n\overline{X}^2)$	$D(\overline{X})=D(X)/n$,	$\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$;
样本标准差	$S=\sqrt{S^2}$		$\overline{X}$ 与 S^2 独立;
样本原点矩	$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$	$E(S^2)=D(X)$.	$\frac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$.
样本中心矩	$B_k=\frac{1}{n\text{-}1}\sum_{i=1}^{n}(X_i-\overline{X})^k$		

三类常见的抽样分布

名称	定义	性质	概率密度图像
χ^2分布	$\sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$ 其中$X_i \sim N(0,1)$	期望(n)、方差$(2n)$、可加性	
t分布	$X\big/\sqrt{Y/n} \sim t(n)$ 其中$X \sim N(0,1)$，$Y \sim \chi^2(n)$ 且独立	极限为N(0,1)、对称性	
F分布	$\dfrac{X/m}{Y/n} \sim F(m,n)$ 其中$X \sim \chi^2(m)$，$Y \sim \chi^2(n)$ 且独立	若$X\sim F(m,n)$，则$1/X\sim F(n,m)$.	







$t_{1-\alpha}(n) = -t_\alpha(n)$　　$F_{1-\alpha}(n_1,n_2) = 1\big/F_\alpha(n_2,n_1)$