西安電子科技大學
XIDIAN UNIVERSITY

第6章 数理统计的基本概念



引言

数理统计是具有广泛应用的一个数学分支，它以概率论为基础，根据试验或观察得到的数据来研究随机现象，对研究对象的客观规律作出种种合理的估计和判断。

数理统计的内容

① 如何收集、整理数据资料；　　数据必须具有随机性

② 如何对所得的数据资料进行分析、研究。

对数据建立一个统计模型，给出统计推断的准则

③ 在给定的统计模式下进行统计推断。

6.1 数理统计的基本概念

1. 总体和样本

总体： 研究对象某个指标取值的全体以及取这些值的可能性。

研究对象的全体元素构成的集合。

总体研究：
1. 指标取哪些值
2. 取这些值的概率

个体： 组成总体的每个元素。

对总体的研究可归结为讨论随机变量 X 的分布（分布函数或分布律或概率密度），及其主要的数字特征（数学期望 EX 和方差 DX）。

样本： 从总体中按一定的规则抽取的一些个体。

规则：要有代表性——每个个体都有可能被抽到，每个个体被抽到的可能性相同

抽样： 抽取样本的行为。

例： 10000件产品，为估计废品率，从中抽取一部分100件进行检查。

总体　　每件产品——个体　　样本　　样本容量大小：100

1. 总体和样本

所抽取的样本必须是随机的，即每一个个体都有同等概率被抽到，而且彼此独立，与总体具有相同的分布。

定义： 称相互独立且与总体 X 同分布的随机变量 $X_1, X_2, \cdots, X_n$ 为来自总体 X 的一个容量为 n 的简单随机样本。

称样本 $X_1, X_2, \cdots, X_n$ 的一组取值 $x_1, x_2, \cdots, x_n$ 为该样本的一组样本值。

2. 统计量

通过观察样本或试验结果的特性对总体特征作出估计和推断是零散的、不系统的，要真正通过样本带来的信息研究总体的特征，就必须对这些信息进行加工、整理，使之系统化，这个过程便引出了统计量的概念。

统计量是为了刻画总体某个特征，对样本的一种加工，即统计量是样本的函数 $g(X_1, X_2, \ldots, X_n)$。

定义：设 $X_1, X_2, \ldots, X_n$ 是来自总体X的一个样本，$g(x_1, x_2, \ldots, x_n)$ 为已知连续函数，且 $g(x_1, x_2, \ldots, x_n)$ 中不含总体 X 的任何未知参数，则称 $g(X_1, X_2, \ldots, X_n)$为总体 X 的一个统计量。

设 $x_1, x_2, \ldots, x_n$ 是样本的一组样本值，则称$g(x_1, x_2, \ldots, x_n)$为统计量的样本值。

设 $X_1, X_2, \ldots, X_n$ 是来自总体X的一个样本，$x_1, x_2, \ldots, x_n$ 是样本 的一组样本值，则可以定义样本均值、样本方差、样本标准差、样本k阶（原点）矩、样本k阶中心矩及其样本值。

2. 统计量

➢几个常用的统计量

样本值:

样本均值

$$\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i;$$

$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i;$$

样本方差

$$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2=\frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^{\,2}-n\bar{X}^2\right)$$

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2=\frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^{\,2}-n\bar{x}^2\right)$$

样本标准差

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

$$s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

样本k阶(原点)矩

$$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k\quad(k=1,2,\ldots)$$

$$a_k=\frac{1}{n}\sum_{i=1}^{n}x_i^k\quad(k=1,2,\ldots)$$

样本k阶中心矩

$$M_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k\quad(k=1,2,\ldots)$$

$$m_k=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^k\quad(k=1,2,\ldots)$$

2. 统计量　➢次序统计量

设 $X_1, X_2, \ldots, X_n$ 是来自总体X的一个样本，将其按大小排列：

$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ ，称（$X_{(1)}, X_{(2)}, \cdots, X_{(n)}$）为次序统计量。

$X_{(1)} = \min(X_1, X_2, \cdots, X_n)$　$X_{(n)} = \max(X_1, X_2, \cdots, X_n)$　统计量。

$X_{(i)}$ 为第i个次序统计量。

6.2 抽样分布

回忆：Γ 函数 $\Gamma(x)=\int_0^{+\infty} t^{x-1}e^{-t}dt \quad x>0$

性质： 1. $\Gamma(x+1)=x\Gamma(x), \Gamma(1)=1 \Rightarrow \Gamma(n+1)=n!$ 2. $\Gamma(\frac{1}{2})=\sqrt{\pi}, \therefore \Gamma(\frac{3}{2})=\frac{1}{2}\Gamma(\frac{1}{2})=\frac{\sqrt{\pi}}{2}$

1. 四大分布

1) 标准正态分布：$X\sim N(0,1)$

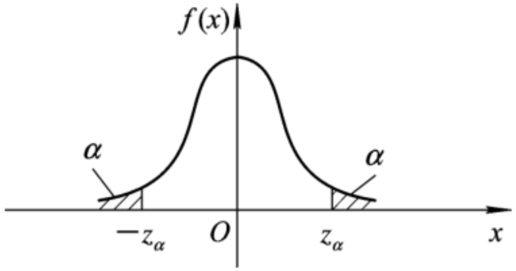
在概率论中，标准正态分布作为重要分布已经作了重点讨论，在此需要补充数理统计中重要的上α分位点的概念。

定义 对于给定的正数 α ($0<\alpha<1$)，如果点 z_α 满足条件

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

则称点 z_α 为*某某分布*的上α分位点。

点 z_α 为$X\sim N(0,1)$的上α分位点。



由于$0<\alpha<1$，因此$0<1-\alpha<1$，从而有上$1-\alpha$分位点，

由标准正态分布的概率密度图的对称性可知 $z_{1-\alpha}=-z_\alpha$

对于$\alpha=0.05$，查得$\Phi(1.645)=0.95$，故$z_{0.05}=1.645$。

对于$\alpha=0.025$，查得$\Phi(1.96)=0.975$，故$z_{0.025}=1.96$。

1. 四大分布

2) χ^2分布 $\chi^2 \sim \chi^2(n)$

定义 设随机变量 相互独立且同服从标准正态分布，则称

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

的分布为服从**自由度为n**的 χ^2分布，记为 $\chi^2 \sim \chi^2(n)$

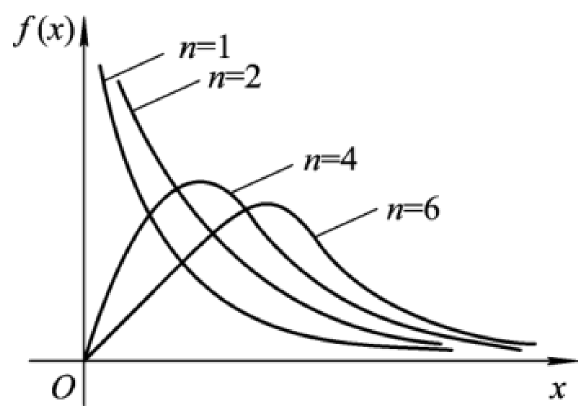
设 $\chi^2 \sim \chi^2(n)$，则χ^2 的概率密度为：

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

当n=1时， $f_1(x) = \dfrac{1}{\Gamma\left(\dfrac{1}{2}\right) 2^{\frac{1}{2}}} x^{-\frac{1}{2}} e^{-\frac{x}{2}}$ ，$x > 0$

当n=2时， $f_2(x) = \dfrac{1}{\Gamma(1)\cdot 2} e^{-\frac{x}{2}}$ ，$x > 0$

当n≥3时，经过原点。



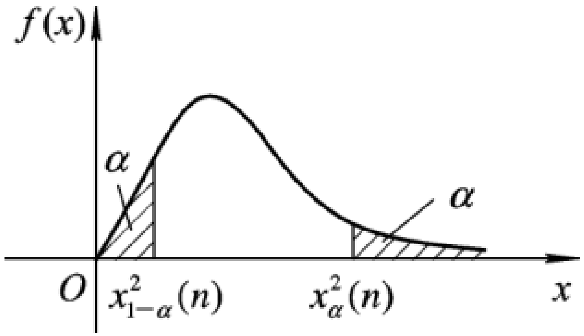
1. 四大分布

2) χ^2分布 $\chi^2 \sim \chi^2(n)$

对于给定的正数α ($0<\alpha<1$)，如果点 $\chi_\alpha^2(n)$ 满足条件

$$P(\chi^2 > \chi_\alpha^2(n)) = \alpha$$

则称点 $\chi_\alpha^2(n)$为 $\chi^2 \sim \chi^2(n)$ 的上 α 分位点。



由于$0<\alpha<1$，因此$0<1-\alpha<1$，从而有上$1-\alpha$分位点 $\chi_{1-\alpha}^2(n)$，由 χ^2 分布的概率密度图不是对称的可知 $\chi_{1-\alpha}^2(n)$ 与 $\chi_\alpha^2(n)$ 没有关系，如图所示。对于不同的 α 和 n，上 α 分位点的值已制成表格（见附表3），可以直接查表。

例如： $\chi_{0.05}^2(10)=18.307$ ， $\chi_{1-0.05}^2(10)=3.94$

1. 四大分布

2) χ^2 分布 $\chi^2 \sim \chi^2(n)$

χ^2分布具有以下性质：

性质 1 设 $\chi^2 \sim \chi^2(n)$，则 $E\chi^2 = n, D\chi^2 = 2n$

证明 设 $\chi^2 \sim \chi^2(n)$，则 $\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2$，其中 $X_1^2 + X_2^2 + \cdots + X_n^2$ 相互独立且同服从标准正态分布$N(0,1)$，从而

$$E\chi^2 = EX_1^2 + EX_2^2 + \cdots + EX_n^2 = nEX_1^2 = n\left[DX_1 + \left(EX_1\right)^2\right] = n$$

$$D\chi^2 = DX_1^2 + DX_2^2 + \cdots + DX_n^2 = nDX_1^2 = n\left[EX_1^4 - \left(EX_1^2\right)^2\right] = n\left(EX_1^4 - 1\right)$$

$$EX_1^4 = \int_{-\infty}^{+\infty} x^4 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^3 d\left(e^{-\frac{x^2}{2}}\right)$$

$$= \left[-\frac{1}{\sqrt{2\pi}} x^3 e^{-\frac{x^2}{2}}\right]_{-\infty}^{+\infty} + 3\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} x^2 e^{-\frac{x^2}{2}} dx = 3EX_1^2 = 3\left[DX_1 + \left(EX_1\right)^2\right] = 3$$

故 $D\chi^2 = 2n$

性质 2 设 $\chi_1^2 \sim \chi^2(n_1), \chi_2^2 \sim \chi^2(n_2)$，且 χ_1^2，χ_2^2 相互独立，则：

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

χ^2分布的可加性

1. 四大分布

3) t 分布：T～$t(n)$

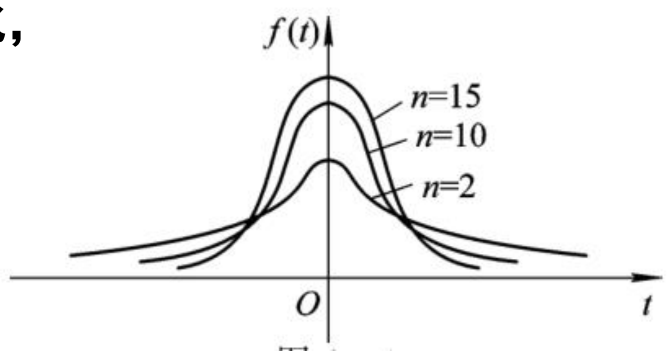
定义 设随机变量X～$N(0,1)$，$Y\sim\chi^2(n)$，且X、Y相互独立，则称 $T=\dfrac{X}{\sqrt{Y/n}}$

的分布为服从参数为n的t分布，记为T～$t(n)$。

设T～$t(n)$，则T的概率密度为：$f(t)=\dfrac{\Gamma\left(\dfrac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\dfrac{n}{2}\right)}\left(1+\dfrac{t^2}{n}\right)^{-\frac{n+1}{2}},-\infty<t<+\infty$

$f(t)$ 的图形如图所示。显然，$f(t)$随n发生变化，且 $f(t)$是偶函数，其图形关于 $t=0$ 对称。**当$n\to\infty$时**，$f(t)$趋于**标准正态分布**的概率密度，即

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

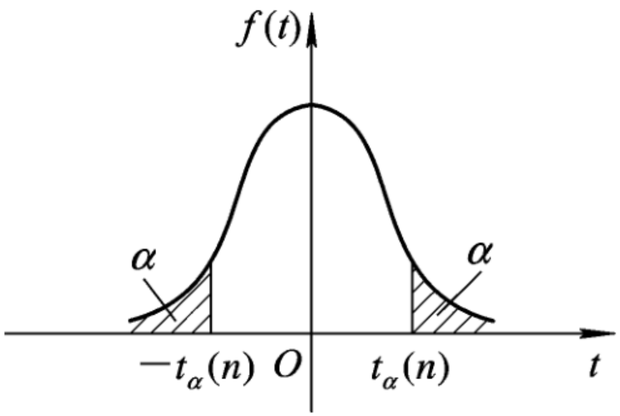


但当n较小时，t分布与标准正态分布差异很大。

1. 四大分布

3) t 分布：$T \sim t(n)$

对于给定的正数α ($0 < \alpha < 1$)，如果点 $t_\alpha(n)$ 满足条件 $P(T > t_\alpha(n)) = \alpha$ 则称点 $t_\alpha(n)$ 为$T \sim t(n)$的上 α 分位点。



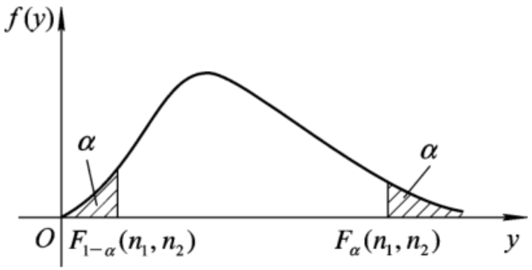
由于$0 < \alpha < 1$，因此$0 < 1-\alpha < 1$，从而有上$1-\alpha$分位点 $t_{1-\alpha}(n)$，由 t 分布的概率密度图的对称性可知 $t_{1-\alpha}(n) = -t_\alpha(n)$ 。

1. 四大分布

4) F分布：$F \sim F(n_1,n_2)$

定义 设随机变量 $X \sim \chi^2(n_1)$，$Y \sim \chi^2(n_2)$，且X、Y相互独立，则称 $F = \dfrac{X/n_1}{Y/n_2}$ 的分布为服从参数为(n_1,n_2)的F分布，记为$F \sim F(n_1,n_2)$。

对于给定的正数α ($0<\alpha<1$)，如果点 $F_\alpha(n_1,n_2)$ 满足条件 $P(F > F_\alpha(n_1,n_2)) = \alpha$，则称点 $F_\alpha(n_1,n_2)$ 为 $F \sim F(n_1,n_2)$的上α分位点。



F分布具有以下性质：

性质 1 设$F \sim F(n_1,n_2)$，则 $\dfrac{1}{F} \sim F(n_2,n_1)$

性质 2 设$F \sim F(n_1,n_2)$，则 $F_{1-\alpha}(n_1,n_2) = \dfrac{1}{F_\alpha(n_2,n_1)}$

性质 3 F分布就是t分布平方： $t^2(n) = \dfrac{X^2}{Y/n} \sim F(1,n)$

2. 正态总体的样本均值和样本方差的分布

设总体X（不管服从什么分布，只要均值和方差存在）的均值为μ，方差为 σ^2， $X_1, X_2, \ldots, X_n$是来自X的一个样本，样本均值 $\overline{X}$，样本方差 S^2。

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

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

$$E(S^2) = E\left[\frac{1}{n-1}\left(\sum_n X_i^2 - n\overline{X}^2\right)\right] = \frac{1}{n-1}\left[\sum_n E(X_i^2) - nE(\overline{X}^2)\right]$$

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

2. 正态总体的样本均值和样本方差的分布

对于正态总体:

定理: （$X_1, X_2, \ldots, X_n$）是从总体 $N(\mu,\sigma^2)$ 中抽取的一个样本，则:

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

② $\overline{X}$ 与 S^2 相互独立

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

2. 正态总体的样本均值和样本方差的分布

③推论：$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$$

证明：根据t分布定义，由于 $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ 与 $\frac{(n-1)S^2}{\sigma^2}$ 独立，

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

$X_i \sim N(\mu,\sigma^2), \frac{X_i-\mu}{\sigma} \sim N(0,1)$ 根据 χ^2分布定义，

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

即：$$\frac{\sum_{n}(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n) \qquad ④$$

2. 正态总体的样本均值和样本方差的分布

对于双正态总体：

双正态总体，设 $X_1,\cdots,X_{n_1}$ 独立且服从 $N(\mu_1,\sigma_1^2)$ ，$Y_1,\cdots,Y_{n_2}$ 独立且服从 $N(\mu_2,\sigma_2^2)$，且样本 $X_1,\cdots,X_{n_1}$ 与 $Y_1,\cdots,Y_{n_2}$ 独立，则：

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

证明：$$\frac{(n_1-1)S_1^2}{\sigma_1^2}\sim\chi^2(n_1-1),\frac{(n_2-1)S_2^2}{\sigma_2^2}\sim\chi^2(n_2-1)$$

且二者相互独立，利用F分布定义：$$\frac{\dfrac{(n_1-1)S_1^2}{\sigma_1^2}\Big/(n_1-1)}{\dfrac{(n_2-1)S_2^2}{\sigma_2^2}\Big/(n_2-1)}\sim F(n_1-1,n_2-1)$$

② 当方差相同时：$\sigma_1^2=\sigma_2^2$

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

其中 $$S_\omega^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

2. 正态总体的样本均值和样本方差的分布

对于双正态总体：

双正态总体，设 $X_1,\cdots,X_{n_1}$ 独立且服从 $N(\mu_1,\sigma_1^2)$ ，$Y_1,\cdots,Y_{n_2}$ 独立且服从 $N(\mu_2,\sigma_2^2)$，且样本 $X_1,\cdots,X_{n_1}$ 与 $Y_1,\cdots,Y_{n_2}$ 独立，则：

③ $$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1) \qquad \overline{X}-\overline{Y} \sim N(\mu_1-\mu_2,\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2})$$

④ $$\frac{n_2\sigma_2^2\sum\limits_{n_1}(X_i-\mu_1)^2}{n_1\sigma_1^2\sum\limits_{n_2}(Y_i-\mu_2)^2} \sim F(n_1,n_2)$$

总结:

单正态总体:

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

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

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

$$\frac{\sum\limits_{n}(X_i-\overline{X})^2}{\sigma^2} \sim \chi^2(n-1)$$

$$\frac{\sum\limits_{n}(X_i-\mu)^2}{\sigma^2} \sim \chi^2(n)$$

双正态总体:

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

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

$$\sigma_1^{\,2}=\sigma_2^{\,2}$$

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

$$\frac{n_2\sigma_2^{\,2}\sum\limits_{n_1}(X_i-\mu_1)^2}{n_1\sigma_1^{\,2}\sum\limits_{n_2}(Y_i-\mu_2)^2} \sim F(n_1,n_2)$$

题型：

1. **求统计量的分布** 已知总体分布（特别是正态分布）求某统计量的分布
2. **利用统计量的分布计算相关事件的概率或与概率相关的问题**
3. **计算统计量的数字特征**

例1： 设总体 $\sim N(0,\sigma^2)$，$X_1,\cdots,X_n$ 为来自总体X的一个样本，则：

$$A、\frac{\overline{X}^2}{\sigma^2} \sim \chi^2(1) \quad B、\frac{S^2}{\sigma^2} \sim \chi^2(n-1) \quad C、\frac{\overline{X}}{S} \sim t(n-1) \quad D、\frac{S^2}{n\overline{X}^2} \sim F(n-1,1)$$

解：

$$\frac{\overline{X}}{\sigma/\sqrt{n}} \sim N(0,1) \qquad \frac{\overline{X}^2}{(\sigma/\sqrt{n})^2} \sim \chi^2(1) \qquad \frac{n\overline{X}^2}{\sigma^2} \sim \chi^2(1)$$

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

$$\frac{\dfrac{(n-1)S^2/\sigma^2}{n-1}}{\dfrac{n\overline{X}^2/\sigma^2}{1}} \sim F(n-1,1) \to \frac{S^2}{n\overline{X}^2} \sim F(n-1,1)$$

例2：$X \sim N(\mu,\sigma^2)$ ， $X_1,\cdots,X_n$ 为来自总体X的一个样本，则：

$$n(\frac{\overline{X}-\mu}{\sigma})^2 \sim \chi^2(1)$$

$$n(\frac{\overline{X}-\mu}{S})^2 \sim F(1,n-1)$$

例3：设二维连续型随机变量（X,Y）的联合概率密度为：

$$f(x,y)=\frac{1}{4\pi}e^{-\frac{1}{4}(x^2+y^2)},-\infty<x<\infty,-\infty<y<\infty$$

则： $\frac{X^2}{Y^2} \sim F(1,1)$

例4：总体 $X \sim N(\mu,\sigma^2)$ ， $X_1,\cdots,X_{16}$ 为来自总体X的一个样本，

若 $P(\bar{X} > \mu + aS) = 0.95$ ，则a= -0. 4383 $(t_{0.05}(15) = 1.7531)$

例5：随机变量X与Y独立，$X \sim N(5,15), Y \sim \chi^2(5)$

求 $P(X-5 > 3.5\sqrt{Y}) =$ 0. 05 $(t_{0.05}(5) = 2.02)$