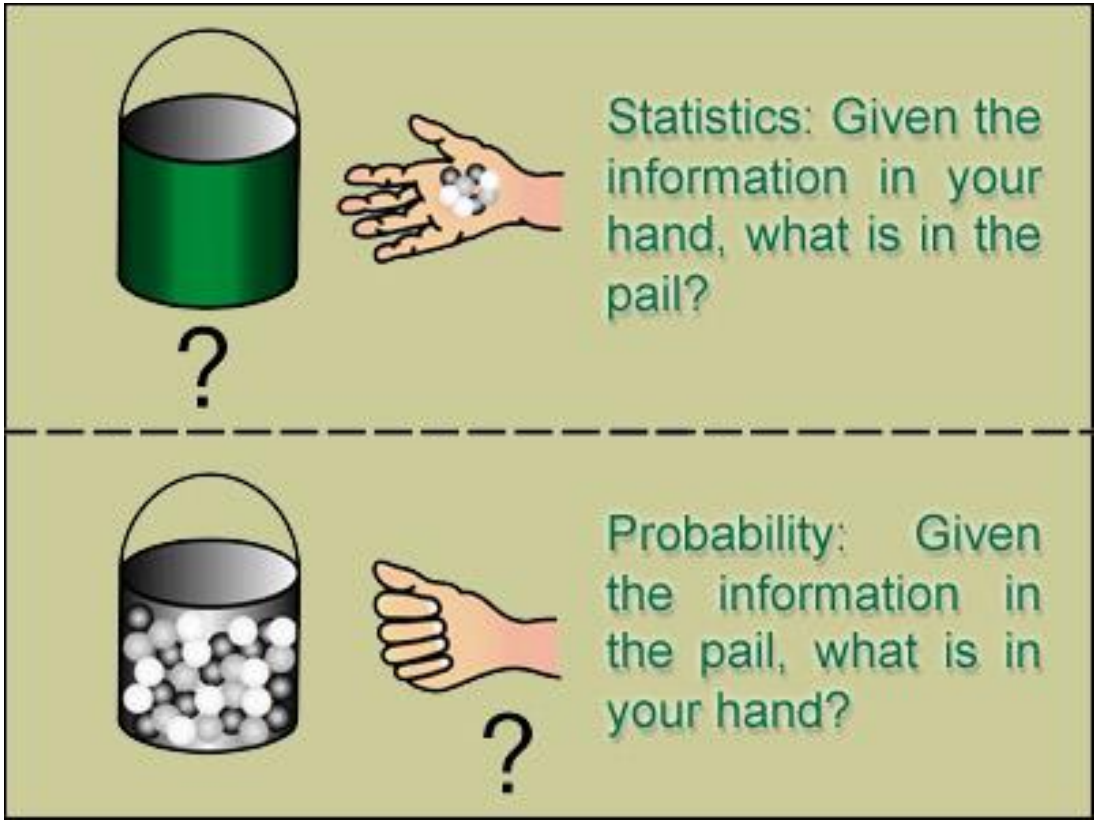
本章转入课程的第二部分

数理统计



Difference between Probability & Statistics

§5.1数理统计学

数理统计学是一门应用性很强的学科. 它是研究怎样以有效的方式收集、整理和分析带有随机性的数据，以便对所考察的问题作出推断和预测，直至为采取一定的决策和行动提供依据和建议.

数理统计不同于一般的资料统计，它更侧重于应用随机现象本身的规律性进行资料的收集、整理和分析.

由于大量随机现象必然呈现出它的规律性，因而从理论上讲，只要对随机现象进行足够多次观察，被研究的随机现象的规律性一定能清楚地呈现出来. 但客观上只允许我们对随机现象进行次数不多的观察试验，也就是说，我们获得的只是局部观察资料.

数理统计的任务就是研究怎样有效地收集、整理、分析所获得的有限的资料，对所研究的问题，尽可能地作出精确而可靠的结论.

数理统计的分类

描述统计学

对随机现象进行观测、试验，以取得有代表性的观测值

推断统计学
Learning in computer science

对已取得的观测值进行整理、分析,作出推断、决策,从而找出所研究的对象的规律性

using data to infer the distribution that generated the data

Statistical inference

Statistical inference, or "**learning**" as it is called in computer science, is the process of *using data to infer the distribution that generated the data*. A typical statistical inference question is:

Given a sample $X_1, \ldots, X_n \sim F$, how do we infer F?

In some cases, we may want to infer only some feature of F such as its mean.

推断统计学

- 参数估计
- 假设检验
- 方差分析
- 回归分析

§ 5.2 总体与样本

总体和样本

总体 —— 研究对象全体元素组成的集合

所研究的对象的某个(或某些)数量指标的全体,它是一个**随机变量**(或多维随机变量).记为X .

X 的分布函数和数字特征称为总体的分布函数和数字特征.

个体 —— 组成总体的每一个元素

即总体的每个数量指标,可看作随机变量 X 的某个取值.用X_i表示.

例如：一个个体X_i指一个人，每个人有2维特征{身高、体重}

样本sample —— 从总体中抽取的部分个体.

用$(X_1,X_2,\ldots,X_n)$表示, n 为样本容量.

称$(x_1,x_2,\ldots,x_n)$为总体 X 的一个容量为n的样本**观测值**,或称**样本的一个实现**.

例如：从全体人中抽取了n个人，及n个人对应的{身高、体重}

样本空间 —— 样本所有可能取值的集合.

例如：所有人，所有的特征可能的取值

总体和样本都是随机变量，观测后才是观测值

简单随机样本

若总体 X 的样本$(X_1,X_2,\ldots,X_n)$满足:

(1) $X_1,X_2,\ldots,X_n$与X 有相同的分布

(2) $X_1,X_2,\ldots,X_n$ 相互独立

则称 $(X_1,X_2,\ldots,X_n)$ 为简单**随机**样本.

一般,对有限总体,放回抽样所得到的样本为简单随机样本,但使用不方便,常用不放回抽样代替.而代替的条件是

$$\underbrace{N}_{\text{总体中个体总数}} / \underbrace{n}_{\text{样本容量}} \geq 10.$$

设总体 X 的分布函数为$F(x)$,则样本 $X_1,X_2,\ldots,X_n$的联合分布函数为

$$F_{总}(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}F(x_i)$$

之间独立、同分布i.i.d.

若总体X 的d.f.为 $f(x)$,则样本的联合 d.f.为

$$f_{总}(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}f(x_i)$$

§5.3 统计量与抽样分布

统计量　　数字特征

定义

设$(X_1,X_2,\dots,X_n)$是取自总体X的一个样本, $g(r_1,r_2,\dots,r_n)$为一实值连续函数,且不含有未知参数,

则称随机变量$g(X_1,X_2,\dots,X_n)$为**统计量**.

若$(x_1,x_2,\dots,x_n)$是一个样本值,

称 $g(x_1,x_2,\dots,x_n)$

为统计量$g(X_1,X_2,\dots,X_n)$的一个**样本值**

例 $X \sim N(\mu,\sigma^2)$，μ,σ^2 是未知参数,

$(X_1,X_2,\dots,X_n)$是一样本, 则

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

是统计量, 其中 $X_i \sim N(\mu,\sigma^2)$

但 $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$ 不是统计量.

若 μ,σ 已知,则为统计量

常用的统计量

重要

设$(X_1, X_2, \ldots, X_n)$ 是来自总体 X 的容量为 n 的样本,称统计量

(1) $$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$ 为**样本均值**

(sample mean)

注意与E(X)的区别

重要

(2) 样本方差 (sample variance)

$$S_n^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2 \text{或} S^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$$

样本标准差

$$S_n=\sqrt{S_n^2} \quad \text{或} \quad S=\sqrt{S^2}$$

重要

(3) $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ 为样本的k阶**原点矩**

(moment)

例如

$$A_1 = \bar{X}$$

$E(X^k)$ —— X的k阶原点矩 (4.3节)

两者的区别？ 分别代表：样本与总体

重要

(4) $B_k=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^k$为样本的$k$阶**中心矩**

例如

$$B_2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\overset{\Delta}{=}S_n^2 \qquad =\frac{n-1}{n}S^2$$

$E((X-E(X))^k)$—总体 X 的 k 阶中心矩

1) $$S^2=\frac{n}{n-1}S_n^2$$

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

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

推导 $$\sum_{i=1}^{n}(X_i-\overline{X})^2=\sum_{i=1}^{n}(X_i^2-2X_i\overline{X}+\overline{X}^2)$$

$$=\sum_{i=1}^{n}X_i^2-2\overline{X}\sum_{i=1}^{n}X_i+\sum_{i=1}^{n}\overline{X}^2=\sum_{i=1}^{n}X_i^2-2n\overline{X}^2+n\overline{X}^2$$

$$=\sum_{i=1}^{n}X_i^2-n\overline{X}^2$$

$$\therefore S_n^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2$$

$$E\left(\overline{X}\right) \qquad D\left(\overline{X}\right) \qquad \mathrm{E}(S^2)$$

2) $E\left(\overline{X}\right)=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=E(X)$ $\qquad$ $D\left(\overline{X}\right)=\frac{1}{n}D(X)$

推导

简单随机抽样的样本

$$E(\bar{X})=E(\frac{1}{n}\sum_{i=1}^{n}X_i)=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=\frac{1}{n}\cdot n\cdot E(X)=\mathrm{E}(X)$$

$$D(\bar{X})=D(\frac{1}{n}\sum_{i=1}^{n}X_i)=\frac{1}{n^2}\sum_{i=1}^{n}D(X_i)=\frac{1}{n^2}\cdot n\cdot D(X)=\frac{1}{n}D(X)$$

$$3)\ E(S_n^2)=\frac{n-1}{n}D(X) \qquad E(S^2)=D(X)$$

推导

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

$$=\frac{1}{n-1}E[X_1^2+X_2^2+X_3^2+\cdots+X_n^2+n\bar{X}^2-2(X_1+X_2+\cdots+X_n)\bar{X}]$$

$$=\frac{1}{n-1}E(X_1^2+X_2^2+X_3^2+\cdots+X_n^2-n\bar{X}^2)$$

$$D(X)=E(X^2)-E^2(X) \qquad E(X^2)=D(X)+E^2(X)=D(X)+E^2(X)=\sigma^2+\mu^2$$

$$E(S^2)=\frac{1}{n-1}\times[n(\sigma^2+\mu^2)-nE(\bar{X}^2)]$$

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

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

抽样分布

统计量既然是依赖于样本的，而后者又是随机变量，故统计量也是随机变量，因而就有一定的分布，这个分布叫做统计量的**抽样分布**.

统计中常用分布

重要

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

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

或 $$Y_n = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \sim N(0,1)$$

中心极限定理

重要

(2) χ^2分布（卡方分布）

χ^2分布是由正态分布派生出来的一种分布.

定义: 设$X_1,X_2,\ldots,X_n$相互独立, 都服从正态分布$N(0,1)$, 则称随机变量:

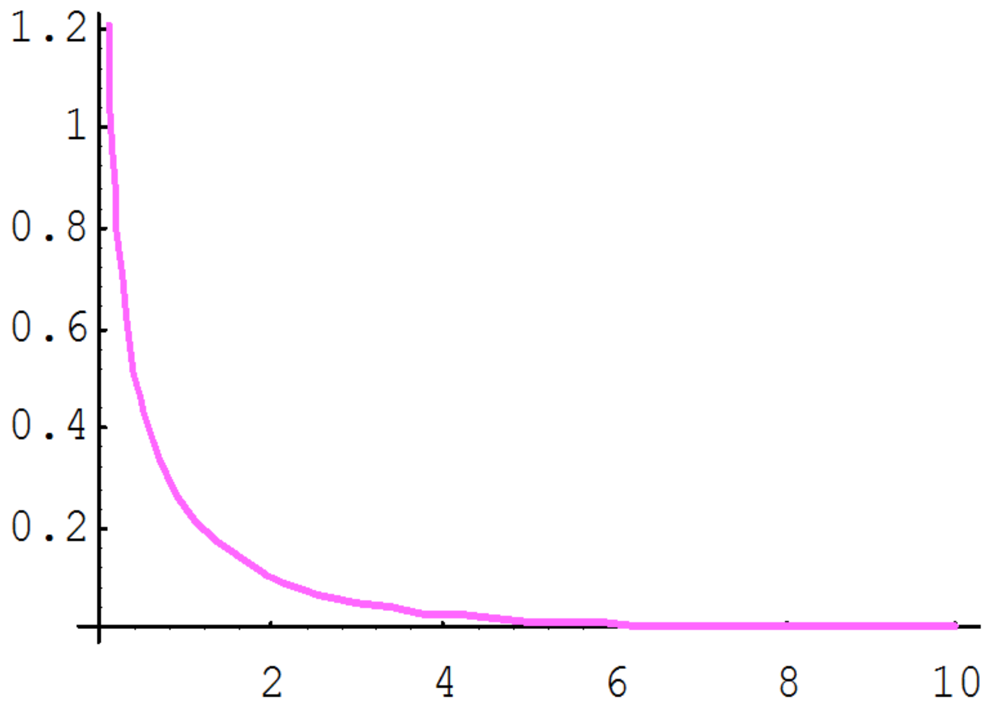
$$\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2$$

所服从的分布为自由度为 n 的χ^2分布.

记为 $\chi^2 \sim \chi^2(n)$

$n=1$ 时,其密度函数为

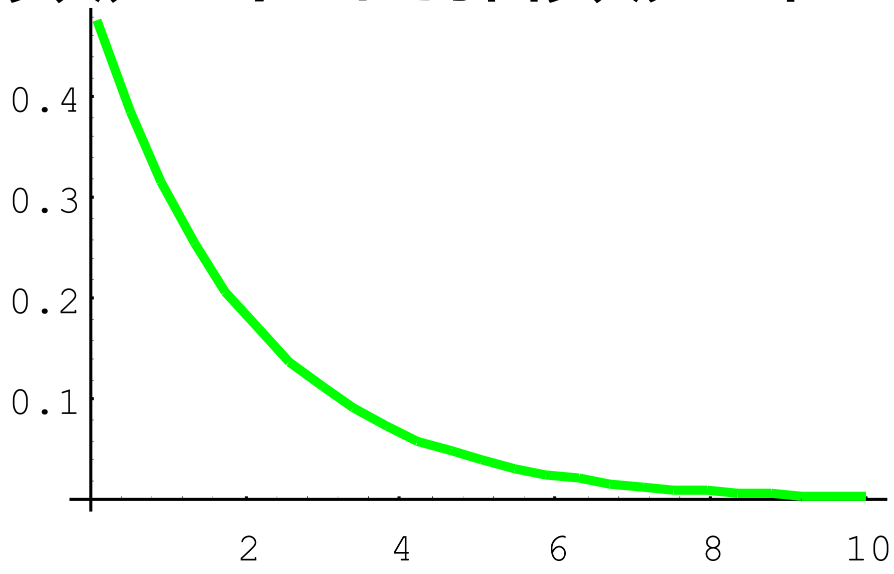
$$f(x)=\begin{cases}\dfrac{1}{\sqrt{2\pi}}x^{-\frac{1}{2}}e^{-\frac{x}{2}}, & x>0\\ 0, & x\le 0\end{cases}$$



$n = 2$ 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.



自由度为 n 的$\chi^2(n)$的密度函数为

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

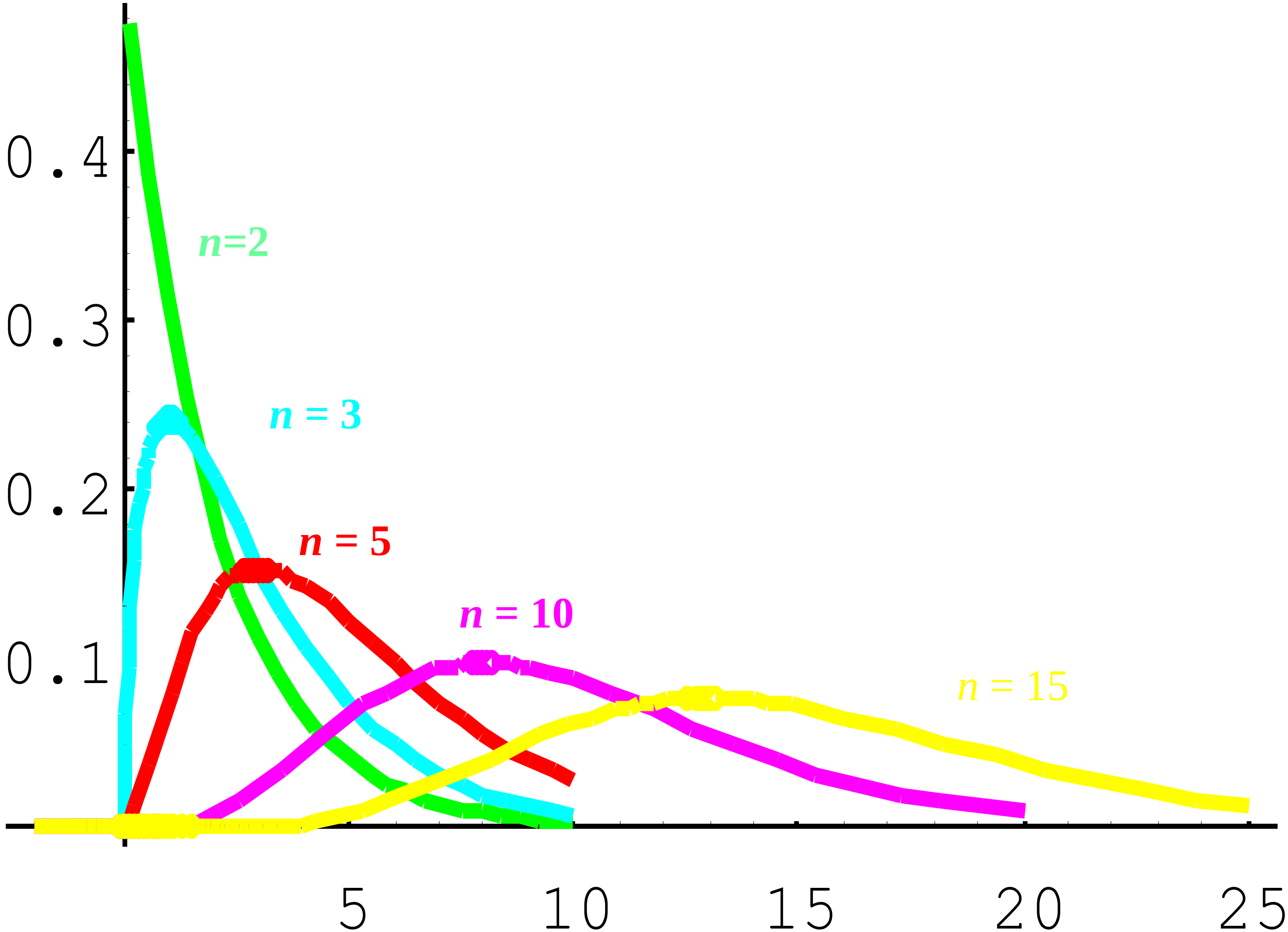
Gamma

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

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

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


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

重要

χ^2分布的性质：

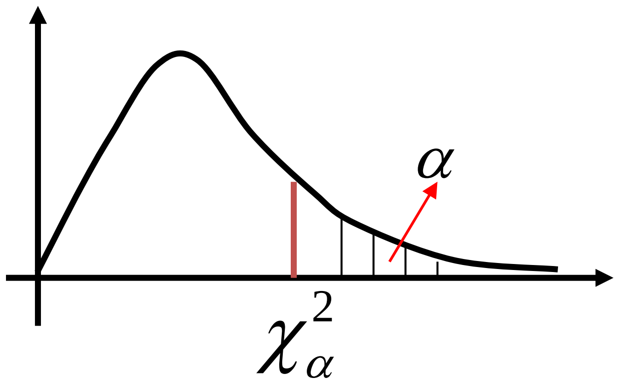
1° $E\left(\chi^2(n)\right)=n, D\left(\chi^2(n)\right)=2n$

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

什么样的正态分布？



对于给定的$\alpha(0<\alpha<1)$，称满足条件：

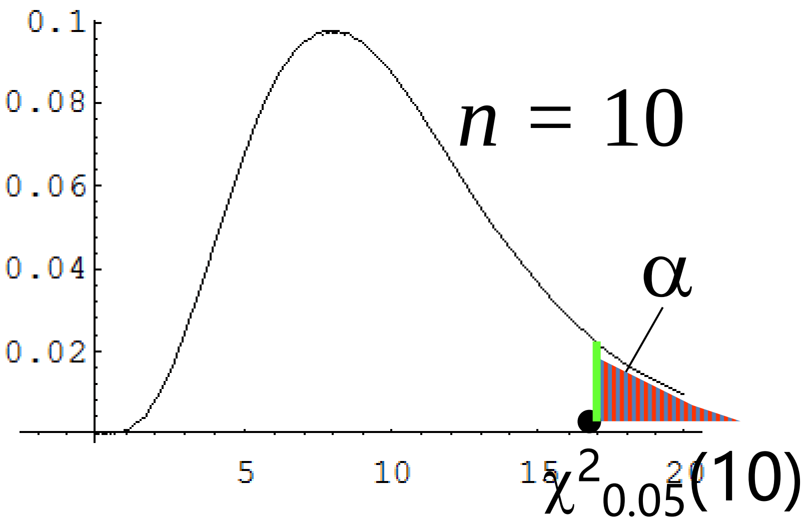
$$P\{\chi^2 > \chi^2_\alpha(n)\} = \alpha$$

的点$\chi^2_\alpha(n)$为$\chi^2(n)$分布的<u>上α分位点</u>。

例如

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

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



(3) t 分布 (Student 分布)

重要

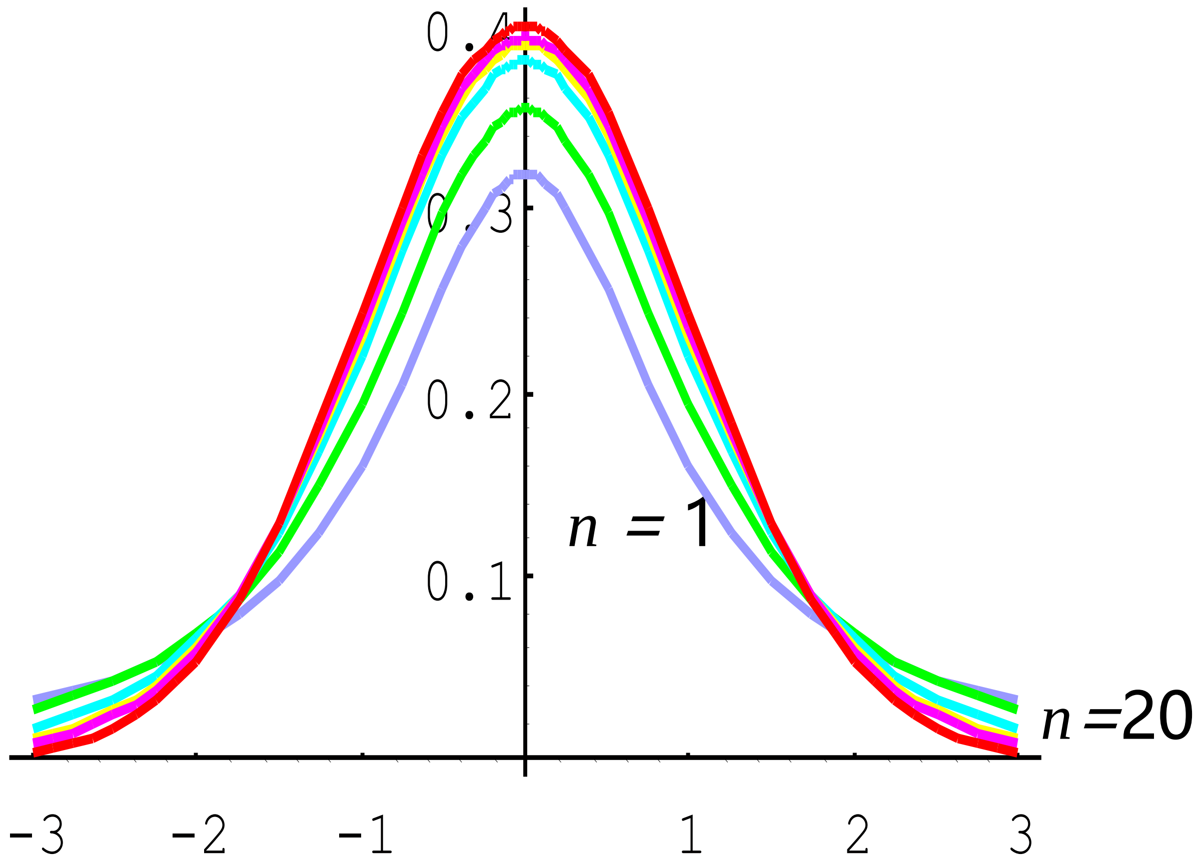
$X \sim N(0,1), Y \sim \chi^2(n), X, Y$独立，则称随机变量

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

为服从自由度是n的$t-$分布，记作$T \sim t(n)$.

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)

演示

t 分布的性质

重要

$f_n(t)$是偶函数,

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

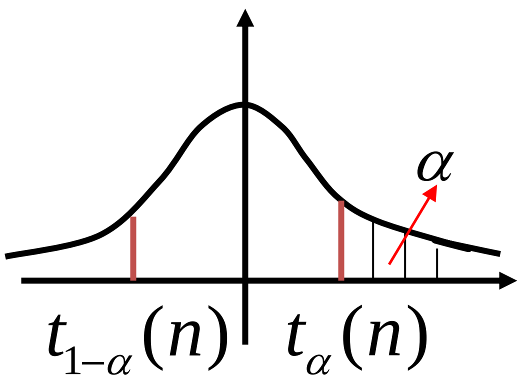
不难看到，当n充分大时，t 分布近似$N(0,1)$分布．但对于较小的n，t分布与$N(0,1)$分布相差很大.

当$n > 30$时，$t_\alpha(n) \approx u_\alpha$.

对于给定的$\alpha(0<\alpha<1)$，称满足条件：

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

的点$t_\alpha(n)$为t分布的上α分位点。



由概率密度的对称性知：$t_{1-\alpha}(n)=-t_\alpha(n)$

(4) F 分布 (Ronald.A.Fisher)

重要

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

则称随机变量

$$F = \frac{X / n_1}{Y / n_2}$$

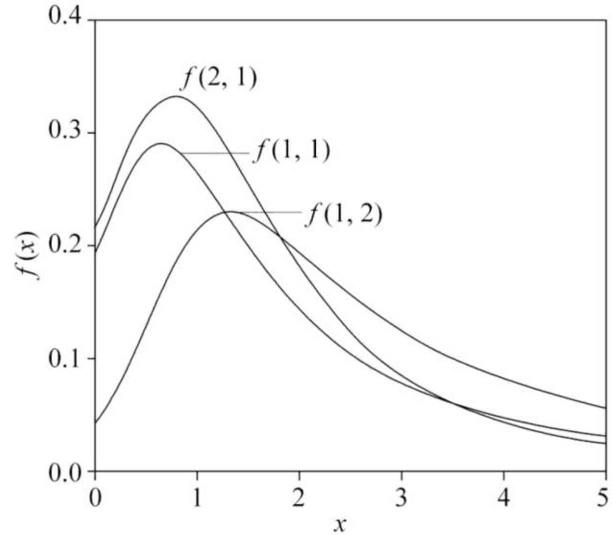
为服从自由度是n_1, n_2 的 $F-$分布，

记作$F \sim F(n_1, n_2)$.

重要

若$F \sim F(n_1,n_2)$， X的概率密度为

$$f(x)=\begin{cases}\dfrac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\frac{n_1}{2})\,\Gamma(\frac{n_2}{2})}\left(\frac{n_1}{n_2}\right)\left(\frac{n_1}{n_2}x\right)^{\frac{n_1}{2}-1}\left(1+\frac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}} & x\ge 0\\ 0 & x<0\end{cases}$$



F 分布的性质

由定义可见，

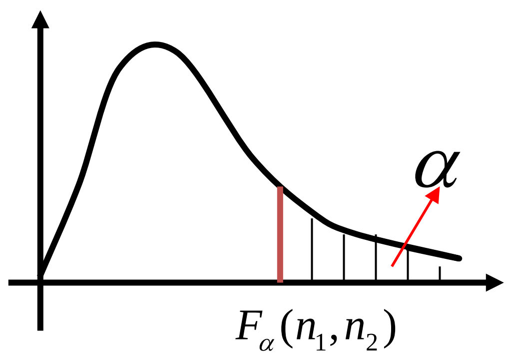
若　$F\sim F(n_1,n_2)$,则　$1/F\sim F(n_2,n_1)$.

演示

对于给定的$\alpha(0<\alpha<1)$，称满足条件：重要

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

的点$F_{\alpha}(n_1,n_2)$为F分布的上α分位点。



结论： $F_{1-\alpha}(n_1,n_2) = 1/F_{\alpha}(n_2,n_1)$

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

证 根据分位点定义

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

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

$$\therefore P\left(\frac{1}{F^{n,m}}\ge\frac{1}{F_{1-\alpha}(n,m)}\right)=\alpha \quad \because \frac{1}{F^{n,m}}\sim F(m,n)$$

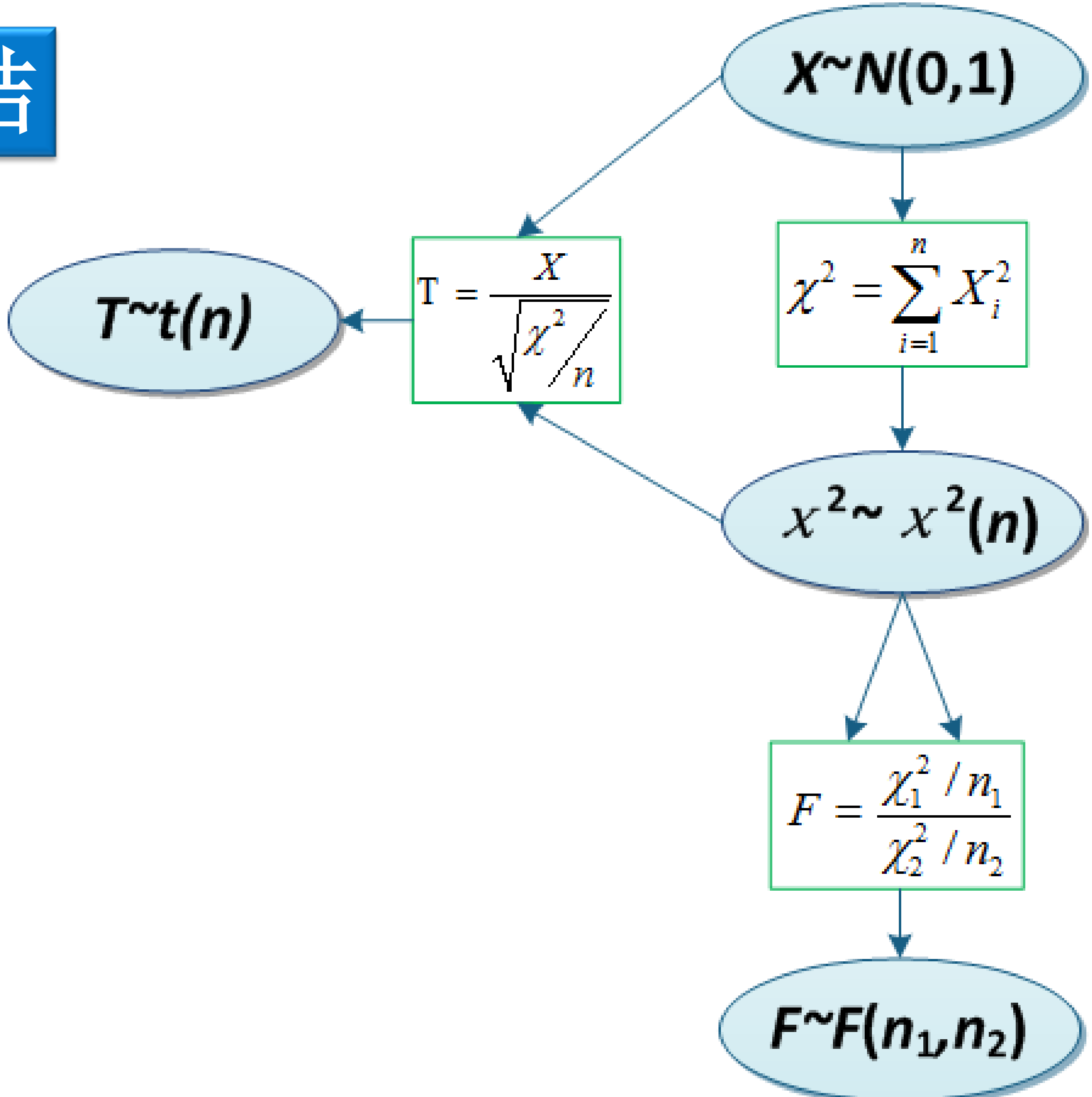
$$\therefore P\left(F^{m,n}\ge\frac{1}{F_{1-\alpha}(n,m)}\right)=\alpha$$

又 $\because F_{\alpha}(m,n)\Rightarrow P(F^{m,n}\ge F_{\alpha}(m,n))=\alpha$

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

总结

重要



$X \sim N(0,1)$
$\chi^2 = \sum_{i=1}^{n} X_i^2$
$T = \frac{X}{\sqrt{\chi^2 / n}}$
$T \sim t(n)$
$\chi^2 \sim \chi^2(n)$
$F = \frac{\chi_1^2 / n_1}{\chi_2^2 / n_2}$
$F \sim F(n_1, n_2)$

正态总体的抽样分布

重要

（I）一个正态总体

定理5.3.1　设$X_1,\cdots,X_n$是总体$N(\mu,\sigma^2)$的样本，

$\bar{X},S^2$分别是样本均值与样本方差，则有：

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

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

(3) $\bar{X}$与S^2独立

(4) $\dfrac{\bar{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$

证明： $(1)\ \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$.

(1) $\because \sum_{i=1}^{n} X_i$ 是相互独立的正态分布的和

$\therefore \sum_{i=1}^{n} X_i$ 也服从正态分布

$$\because E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i) = n\mu \quad D(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} D(X_i) = n\sigma^2$$

$$\therefore \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$$

$$\therefore \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \frac{\sigma^2}{n})$$

将 $\bar{X}$ 标准化 $U = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

证明：(2) $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

（2）令 $Z_i = \dfrac{X_i-\mu}{\sigma}(i=1,2,\cdots,n)$

则$Z_1, Z_2, \cdots, Z_n$相互独立，且都服从$N(0,1)$。从而 $\overline{Z} = \dfrac{1}{n}\sum_{i=1}^{n} Z_i = \dfrac{\overline{X}-\mu}{\sigma}$

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

$$= \sum_{i=1}^{n}\left(Z_i-\overline{Z}\right)^2$$

$$= \sum_{i=1}^{n} Z_i^2 - n\overline{Z}^2$$

取一n阶正交矩阵$A=(a_{ij})_{n*n}$，其中第n行元素均为$\dfrac{1}{\sqrt{n}}$，作正交变换 $Y=AZ$

其中$Y=(Y_1,Y_2,\cdots,Y_n)^{\mathrm{T}}$，$Z=(Z_1,Z_2,\cdots,Z_n)^{\mathrm{T}}$。

由于 $Y_i=\sum_{j=1}^{n}a_{ij}Z_j\left(i=1,2,\cdots,n\right)$，故$Y_1,Y_2,\cdots,Y_n$也服从正态分布。

由$Z\sim N_n(0,E_n)$知，

Cov(AX,BY)=ACov(X,Y)B^{T}

$$E(Y)=E(AZ)=AE(Z)=0$$

$$D(Y)=\mathrm{Cov}(Y,Y)=\mathrm{Cov}(AZ,AZ)=A\mathrm{Cov}(Z,Z)A^T=AE_nA^T=E_n$$

由正交矩阵的性质，故$Y_1,Y_2,\cdots,Y_n$两两不相关，而且

$Y_i\sim N(0,1)(i=1,2,\ldots,n)$。

Cov为单位矩阵，相关系数为0

又由于$Y=(Y_1,Y_2,\cdots,Y_n)^{\mathrm{T}}$是由$Z=(Z_1,Z_2,\cdots,Z_n)^{\mathrm{T}}$经过线性变换而得到的，于是由$Y_1,Y_2,\cdots,Y_n$两两不相关（+正态）可推得$Y_1,Y_2,\cdots,Y_n$相互独立。

而 $$Y_n=\sum_{j=1}^{n}a_{nj}Z_j=\sum_{j=1}^{n}\frac{1}{\sqrt{n}}Z_j=\sqrt{n}\,\overline{Z}$$

$$\sum_{i=1}^{n}Y_i^2=Y^TY=(AZ)^T(AZ)=Z^T\left(A^TA\right)Z=Z^TE_nZ=Z^TZ=\sum_{i=1}^{n}Z_i^2$$

于是 $$\frac{(n-1)S^2}{\sigma^2}=\sum_{i=1}^{n}Z_i^2-n\overline{Z}^2=\sum_{i=1}^{n}Y_i^2-Y_n^2=\sum_{i=1}^{n-1}Y_i^2$$

由于$Y_1,Y_2,\cdots,Y_{n-1}$相互独立，且$Y_i \sim N(0,1)(i=1,2,\ldots,n-1)$，

$$\therefore \sum_{i=1}^{n-1}Y_i^2\sim\chi^2(n-1)$$

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

证明： (3) $\bar{X}$与S^2独立

（3）$\because\ \overline{X} = \sigma \overline{Z} + \mu = \dfrac{\sigma}{\sqrt{n}} Y_n + \mu$ 只依赖于Y_n

而 $S^2 = \dfrac{\sigma^2}{n-1} \sum_{i=1}^{n-1} Y_i^2$ 仅依赖于$Y_1, Y_2, \cdots, Y_{n-1}$

$\therefore\ \overline{X}$与S^2相互独立

证明： (4) $\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

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

而且两者相互独立，根据t分布的定义得

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

（II）两个正态总体

定理2

设$X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$, 且$X$与$Y$独立, $X_1, X_2, \ldots, X_{n_1}$是取自$X$的样本, $Y_1, Y_2, \ldots, Y_{n_2}$是取自$Y$的样本, $\overline{X}$和$\overline{Y}$分别是这两个样本的样本均值，S_1^2和S_2^2分别是这两个样本的样本方差, 则有

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

$$\sigma_1^2 = \sigma_2^2$$

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

5.3.1设X_1, X_2, X_3, X_4是来自正态总体$N(0,9)$的样本，$Y=a(X_1-X_2)^2+b(3X_3+4X_4)^2$,问当$a,b$各为多少时，统计量$Y$服从$\chi^2(2)$分布？

$$X_1-X_2 \sim N(0,2*3^2),\text{即} \frac{X_1-X_2}{3\sqrt{2}} \sim N(0,1)$$

$$3X_3+4X_4 \sim N(0,25*3^2),\text{即} \frac{3X_3+4X_4}{15} \sim N(0,1)$$

根据$\chi^2(2)$分布定义

$$Y=\left(\frac{X_1-X_2}{3\sqrt{2}}\right)^2+\left(\frac{3X_3+4X_4}{15}\right)^2 \sim \chi^2(2)$$

$$Y=\frac{1}{18}(X_1-X_2)^2+\frac{1}{225}(3X_3+4X_4)^2$$

$$\therefore a=\frac{1}{18},\ \text{b}=\frac{1}{225}$$

5.3.2设随机变量X和Y相互独立且都服从正态分布$N(0,9)$，而$X_1,X_2,\dots X_9$和$Y_1,Y_2,\dots Y_9$分别是来自总体X和Y的简单随机样本，则统计量U服从什么分布？$U=\dfrac{X_1+X_2+\dots+X_9}{\sqrt{Y_1^2+Y_2^2+\dots+Y_9^2}}$

令$W_i=\dfrac{X_i}{3},S_i=\dfrac{Y_i}{3},i=1,2,\dots,9,$则

$$W_i\sim N(0,1),S_i\sim N(0,1)$$

$$W=W_1+W_2+\dots+W_9\sim N(0,3^2)$$

$$S=S_1^2+S_2^2+\dots+S_9^2\sim\chi^2(9)$$

$$\therefore U=\frac{X_1+X_2+\dots+X_9}{\sqrt{Y_1^2+Y_2^2+\dots+Y_9^2}}=\frac{W_1+W_2+\dots+W_9}{\sqrt{S_1^2+S_2^2+\dots+S_9^2}}=\frac{W}{\sqrt{S}}=\frac{W/3}{\sqrt{S/9}}$$

$$\because W/3\sim N(0,1),S\sim\chi^2\quad\therefore U\sim t(9)$$

5.3.3设总体X~N(0,2²)，$X_1,X_2,\ldots,X_{15}$是来自总体X的样本，则统计量

$$Y=\frac{2(X_1^2+X_2^2+\ldots+X_5^2)}{X_6^2+X_7^2+\ldots+X_{15}^2}$$

服从什么分布？

$$\because S_1^2=\left(\frac{X_1}{2}\right)^2+\left(\frac{X_2}{2}\right)^2+\ldots+\left(\frac{X_5}{2}\right)^2\sim\chi^2(5)$$

$$S_2^2=\left(\frac{X_6}{2}\right)^2+\left(\frac{X_7}{2}\right)^2+\ldots+\left(\frac{X_{15}}{2}\right)^2\sim\chi^2(10)$$

$$\therefore Y=\frac{2(X_1^2+X_2^2+\ldots+X_5^2)}{X_6^2+X_7^2+\ldots+X_{15}^2}=\frac{2S_1^2}{S_2^2}=\frac{S_1^2/5}{S_2^2/10}\sim F(5,10)$$