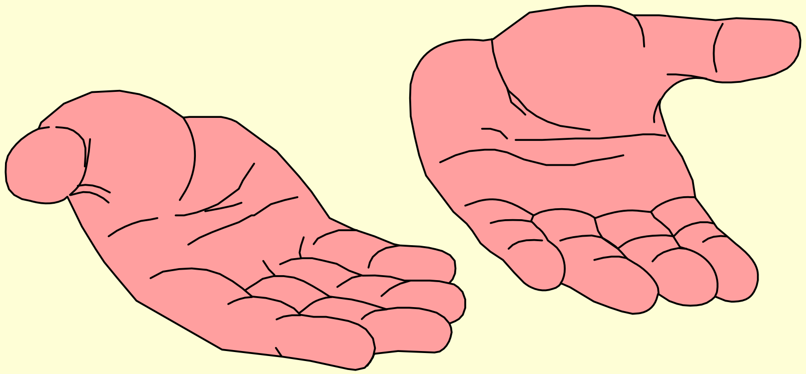
第六章 数理统计的基本概念

- 总体与样本
- 常用的统计量分布
- 抽样分布



引 言

数理统计学是数学的一个重要分支，它研究怎样有效地收集、整理和分析带有随机性的数据，以对所考察的问题作出推断或预测，并为采取一定的决策和行动提供依据和建议。

几个实际问题：

1. 估计产品寿命问题: 根据用户调查获得某品牌洗衣机**50**台的使用寿命为，**5**，**5.5**，**3.5**，**6.2**，**……..**。根据这些数据希望得到如下推断：

A．可否认为产品的平均寿命不低于**4**年？

B．保质期设为多少年，才能保证有**95%**以上的产品过关？

2．商品日投放量问题：如草莓的日投放量多少合理？如何安排银行各营业网点的现金投放量？快餐食品以什么样的速度生产最为合理等等。

例 制衣厂为了合理的确定服装各种尺码的生产比例，需要调查人们身长的分布。现从男性成人人群中随机选取100人，得到他们的身长数据为： …

（1）试推断男性成人身长X的概率密度

(2)若已知X服从正态分布$N(\mu, \sigma^2)$，

试估计参数的μ, σ^2值

已知“总体”的分布类型,对分布中的未知参数所进行的统计推断属于“参数统计”.

§1 总体与样本

1. 总体：研究对象的全体。

通常指研究对象的某项数量指标。

组成总体的元素称为个体。

从本质上讲，总体就是所研究的随机变量或随机变量的分布。

2. 样本 来自总体的部分个体X_1，…，X_n
如果满足：

（1）同分布性： X_i，$i=1,\ldots,n$与总体同分布.

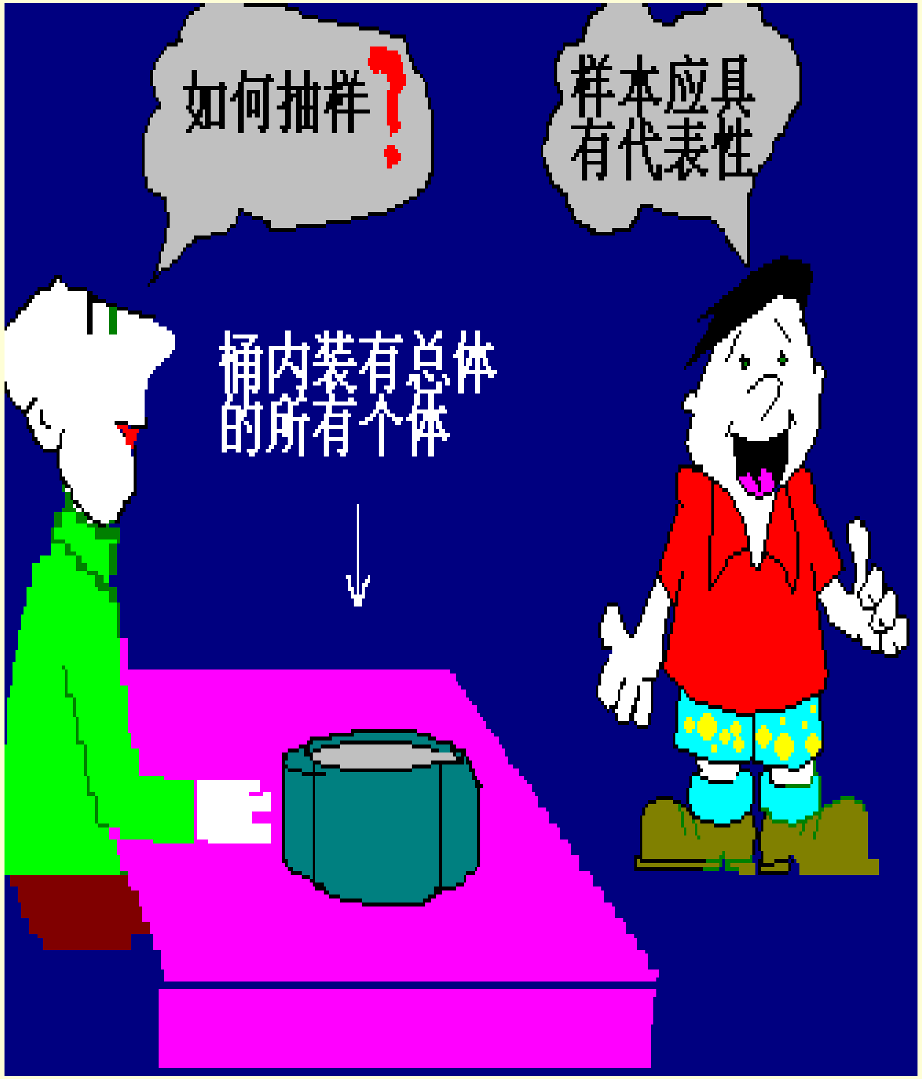
（2）独立性：

X_1，…，X_n 相互独立；

则称为容量为n 的简单随机样本，简称样本。

而称X_1，…，X_n 的一次实现为样本观察值，记为

x_1，…，x_n



来自总体X的随机样本X_1，…，X_n可记为

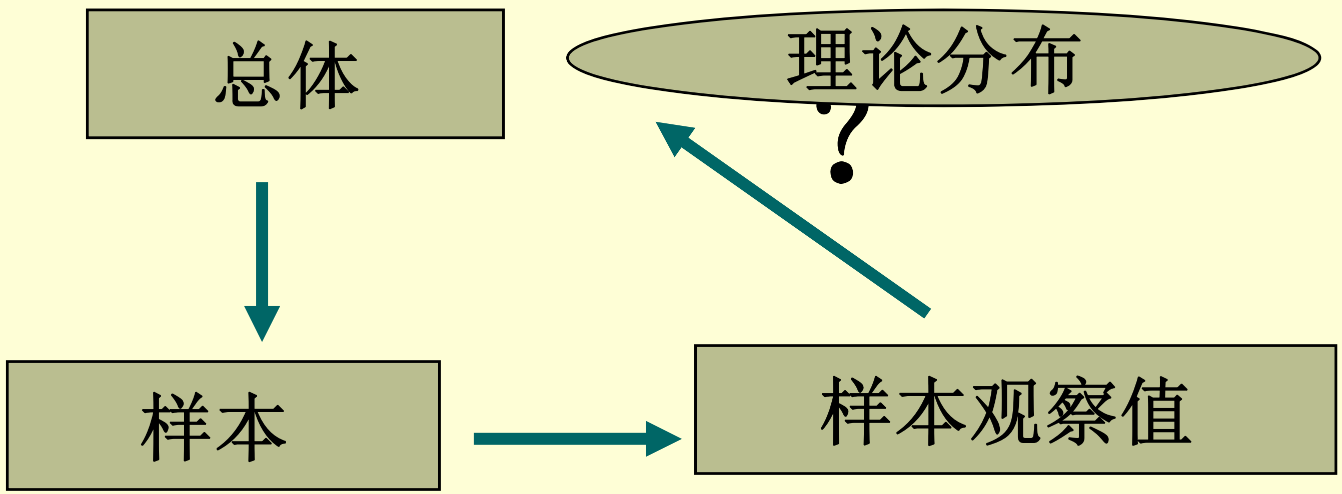
$$X_1,\cdots,X_n \overset{iid}{\sim} X \text{ 或} f(x), F(x),\ldots$$

显然，样本联合分布函数或密度函数为

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

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

3.总体、样本、样本观察值的关系



统计是从手中已有的资料——样本观察值，去推断总体的情况——总体分布。样本是联系两者的桥梁。总体分布决定了样本取值的概率规律，也就是样本取到样本观察值的规律，因而可以用样本观察值去推断总体

统计量

定义： 称样本X_1， …， X_n 的函数

$g(X_1,\ \dots,\ X_n)$是总体X的一个统计量，如果

$g(X_1,\ \dots,\ X_n)$不含未知参数

几个常用的统计量 ：

1. 样本均值 $\overline{X}=\dfrac{1}{n}\sum_{i=1}^{n}X_i,$

2. 样本方差 $S^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$

样本均方差(标准差) $S=\sqrt{S^2},$

3.样本k阶矩

样本原点矩 $$\frac{1}{n}\sum_{i=1}^{n}X_i^k$$

样本中心矩 $$\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k,$$

4.经验分布函数 用S(x)表示样本X_1， …， X_n中不大于x的随机变量个数。定义经验分布函数$F_n(x)$为

$$F_n(x)=\frac{1}{n}S(x)$$

§2 常用统计量的分布

统计量的分布称为抽样分布。数理统计中常用到如下三个分布：

χ^2分布、t 分布和F分布。

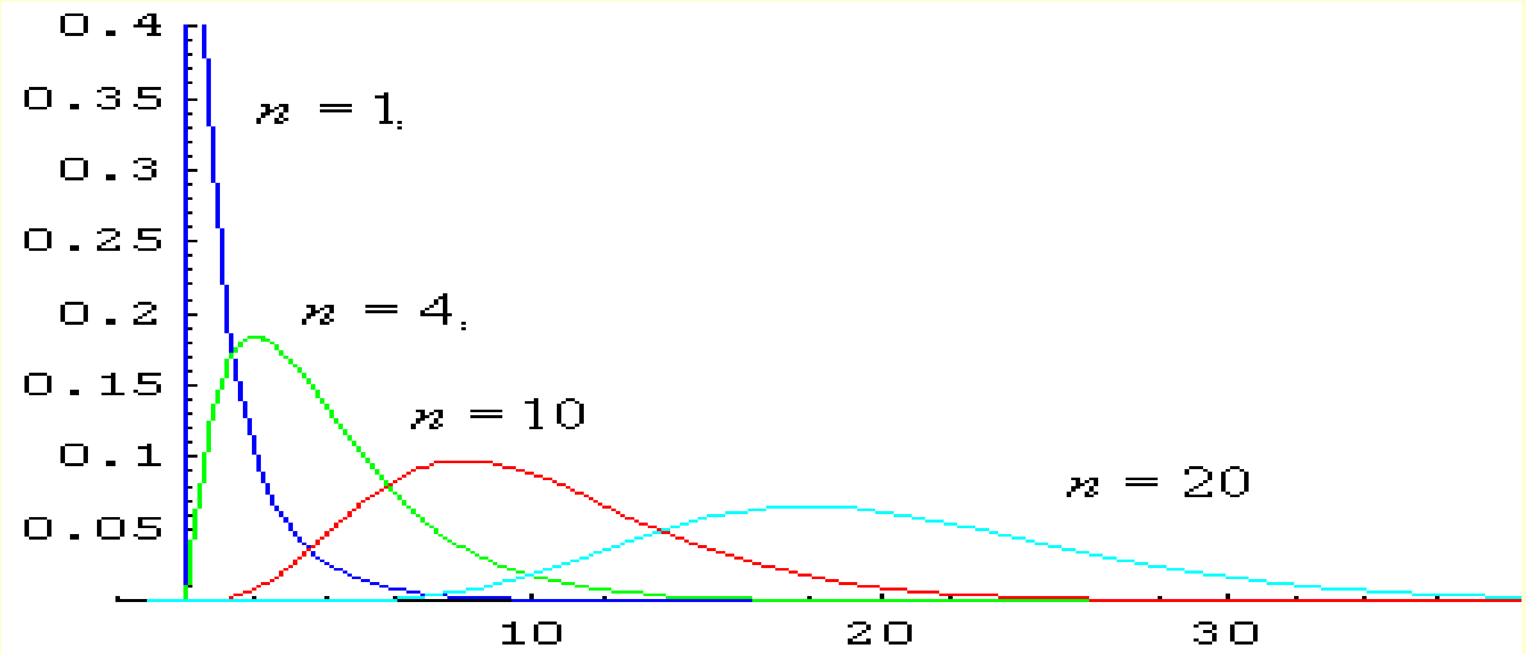
一、 χ^2分布

1. 构造　设 $X_1,\cdots,X_n \overset{iid}{\sim} N(0,1)$，　则$\chi^2=\sum_{i=1}^{n}X_i^2 \sim \chi^2(n)$.

称为自由度为n的χ^2分布.

2.χ^2分布的密度函数**f(y)**曲线

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

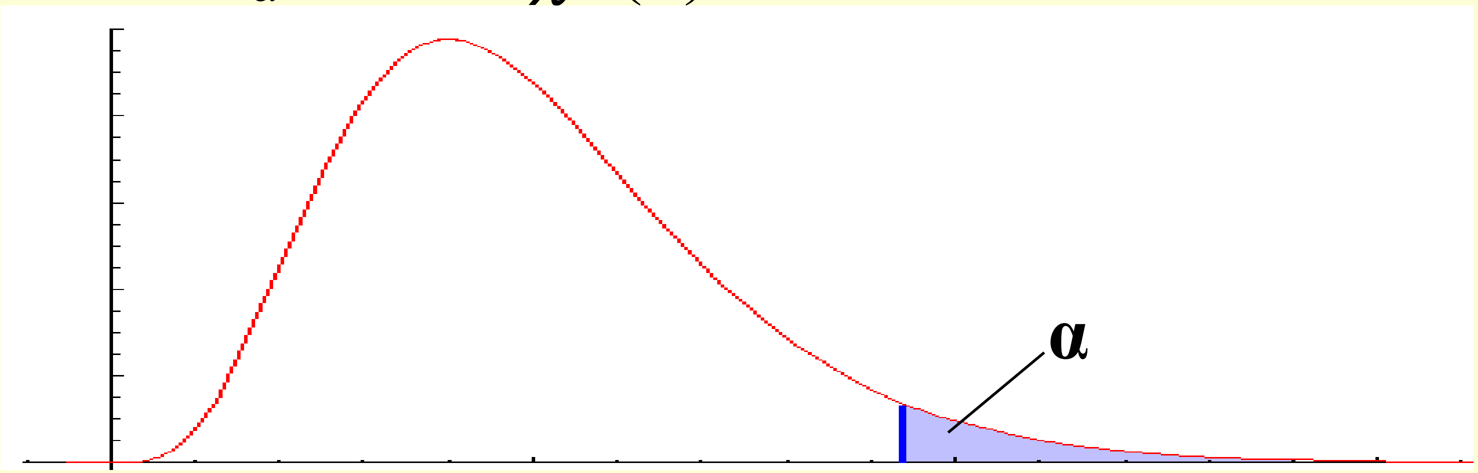


3. 分位点 设**X ~ χ²(n)**，若对于α：**0<α<1**，

存在 $\chi_{\alpha}^{2}(n)>0$

满足 $P\{X \ge \chi_{\alpha}^{2}(n)\}=\alpha,$

则称 $\chi_{\alpha}^{2}(n)$ 为 $\chi^{2}(n)$ 分布的上α分位点。



$\chi_{\alpha}^{2}(n)$

如： $\chi_{0.05}^{2}(10)=18.307$ $\chi_{0.95}^{2}(10)=3.940$

4.性质：

a.分布可加性 若$X \sim \chi^2(n_1)$，$Y \sim \chi^2(n_2)$， X，Y独立，则 $X+Y \sim \chi^2(n_1+n_2)$

b.期望与方差 若$X \sim \chi^2(n)$，则

$$E(X)=n，D(X)=2n$$

二、t分布

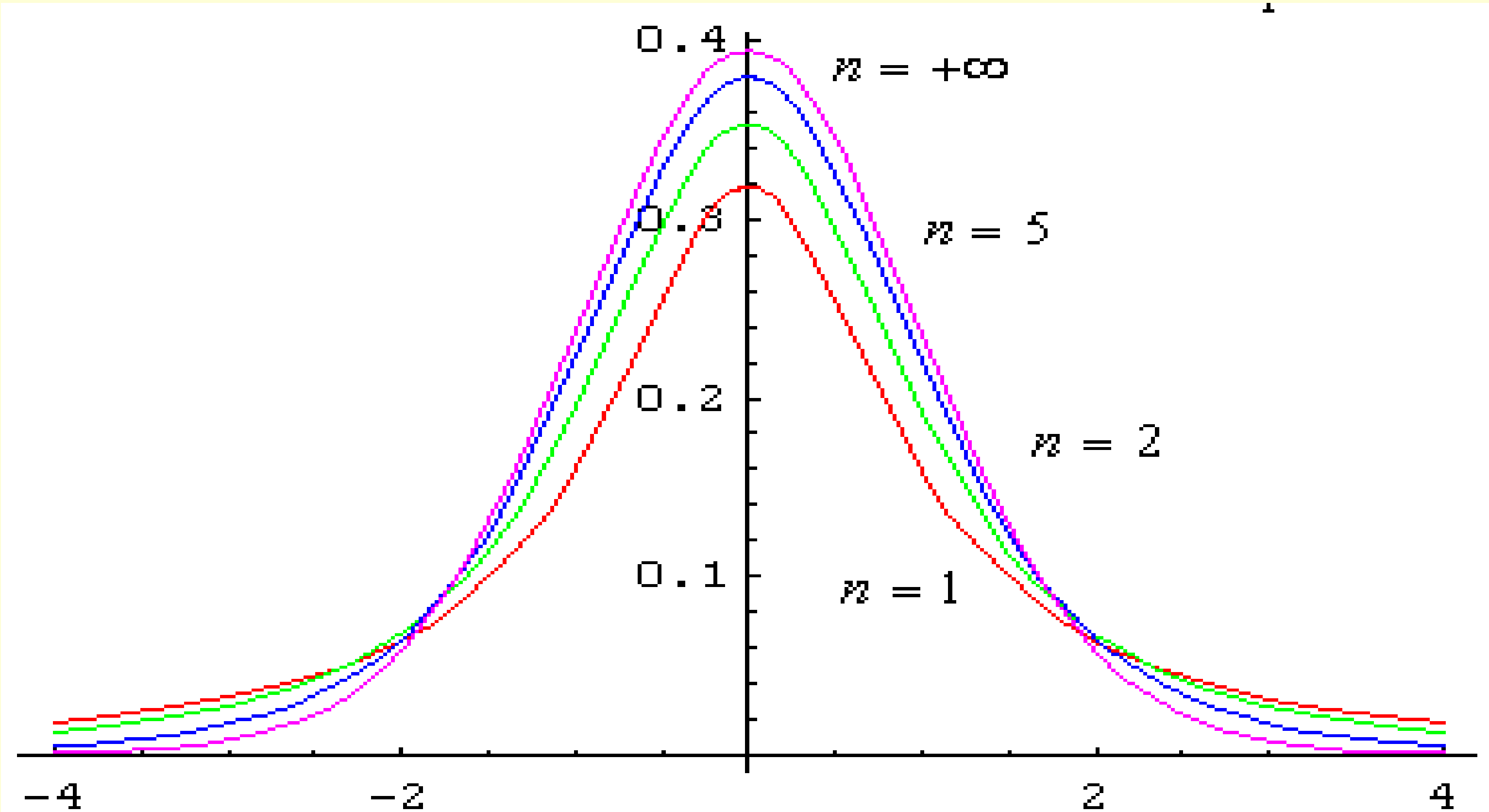
1.构造 若$\xi \sim N(0,1)$, $\eta \sim \chi^2(n)$, ξ与η独立，则

$$T=\frac{\xi}{\sqrt{\eta/n}} \sim t(n).$$

t(n)称为自由度为**n**的**t**分布。

t(n) 的概率密度为

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



2. 基本性质：

(1) f(t)关于t=0(纵轴)对称。

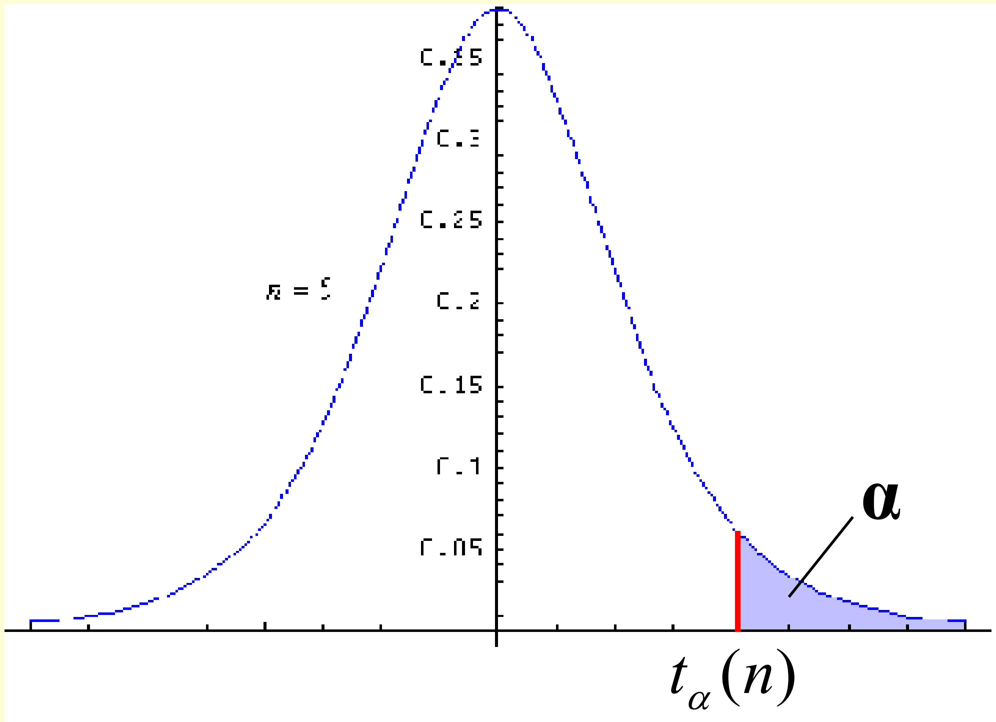
(2) f(t)的极限为N(0，1)的密度函数，即

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

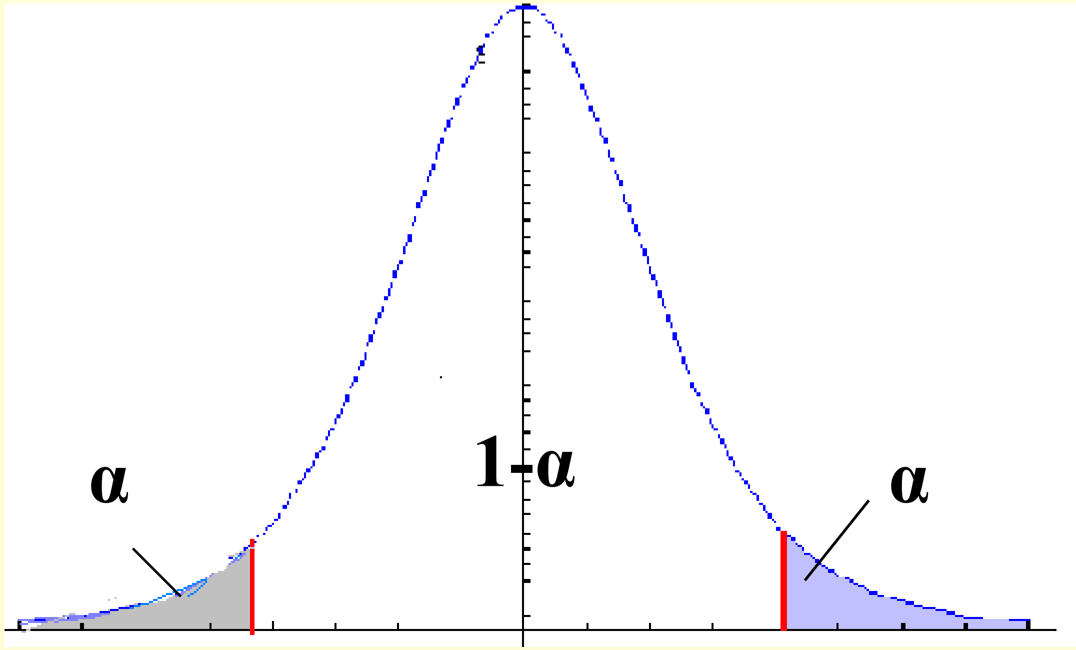
3. 分位点

设T～t(n)，若对$\alpha: 0<\alpha<1$，存在$t_\alpha(n)$，满足$P\{T \geq t_\alpha(n)\} = \alpha$，

则称$t_\alpha(n)$为t(n)的上侧α分位点



注： $t_{1-\alpha}(n) = -t_{\alpha}(n)$



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

如： $t_{0.01}(20) = 2.5280 \quad t_{0.01/2}(20) = 2.8453$

$$t_{0.99}(20) = t_{1-0.01}(20) = -t_{0.01}(20) = -2.5280$$

三、F分布

1.构造 若$\eta_1 \sim \chi^2(n_1)$， $\eta_2 \sim \chi^2(n_2)$，η_1， η_2独立，则

$$F = \frac{\eta_1 / n_1}{\eta_2 / n_2} \sim F(n_1, n_2).$$

称为第一自由度为n_1，第二自由度为n_2的F分布,其概率密度为

$$h(y) = \begin{cases} \dfrac{\Gamma(\frac{n_1+n_2}{2})(n_1/n_2)^{n_1/2} y^{\frac{n_1}{2}-1}}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})(1+\frac{n_1}{n_2}y)^{(n_1+n_2)/2}}, & y > 0 \\ 0, & y \le 0 \end{cases}$$

2. F分布的分位点

对于α：0<α<1，

若存在$F_\alpha(n_1, n_2)>0$，

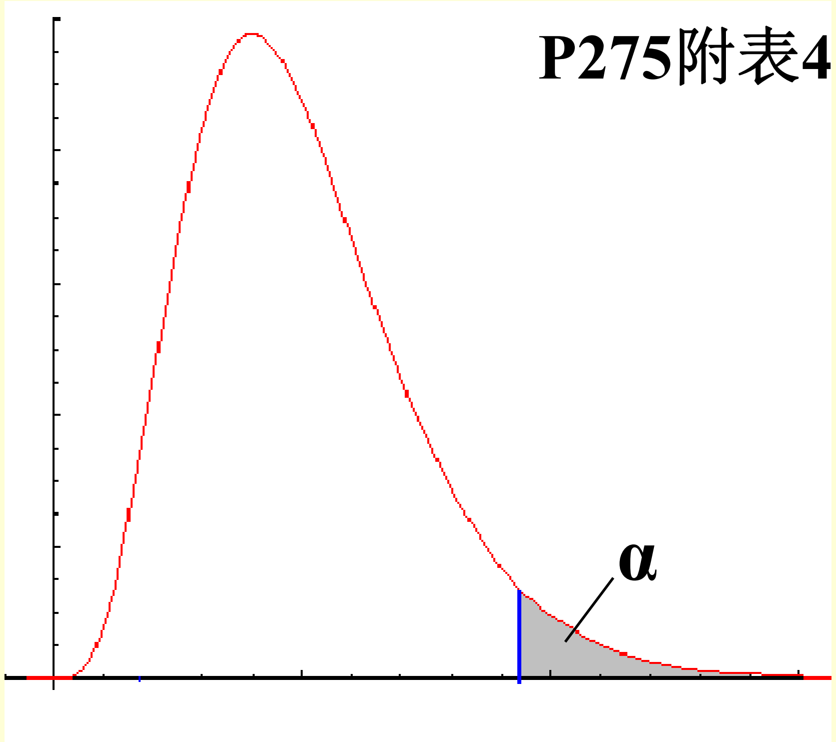
满足

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

称$F_\alpha(n_1, n_2)$为

$F(n_1, n_2)$的

上侧α分位点；



$F_\alpha(n_1, n_2)$

注：
$$F_{1-\alpha}(n_1,n_2)=\frac{1}{F_\alpha(n_2,n_1)}$$

证明： 设F~F(n$_1$,n$_2$)，则 $\frac{1}{F}\sim F(n_2,n_1)$

$$P\{F\geq F_{1-\alpha}(n_1,n_2)\}=1-\alpha$$

$$P\{\frac{1}{F}>F_\alpha(n_2,n_1)\}=\alpha$$

$$P\{\frac{1}{F}\leq\frac{1}{F_{1-\alpha}(n_1,n_2)}\}=1-\alpha$$

$$P\{\frac{1}{F}>\frac{1}{F_{1-\alpha}(n_1,n_2)}\}=\alpha$$

得证!

§3 抽样分布

1.若$X_1,\cdots,X_n \overset{iid}{\sim} N(\mu,\sigma^2)$, 则$U=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$

证明: $\overline{X}=\dfrac{1}{n}\sum_{i=1}^{n}X_i \quad E(\overline{X})=\dfrac{1}{n}\sum_{i=1}^{n}E(X_i)=\mu$

是n 个独立的正态随机变量的线性组合,故服从正态分布

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

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

2.若$X_1,\cdots,X_n \overset{iid}{\sim} N(\mu,\sigma^2)$，则

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

(3) $T=\dfrac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$.

(1),(2)证略。以下证明(3)：

$$\because U=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)\qquad V=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1);$$

且U与V独立,根据t分布的构造 $\dfrac{U}{\sqrt{\dfrac{V}{n-1}}}\sim t(n-1)$

得证!

3.若$X_1,\cdots,X_{n_1} \overset{iid}{\sim} N(\mu_1,\sigma_1^2), Y_1,\cdots,Y_{n_2} \overset{iid}{\sim} N(\mu_2,\sigma_2^2)$,

且两样本独立.则

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

(2)进一步,假定$\sigma_1^2=\sigma_2^2$,就有,

$$T=\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{S_w\sqrt{1/n_1+1/n_2}}\sim t(n_1-1+n_2-1).$$ 其中

$$S_w^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$称为混合样本方差.

证明:(1) $\because \dfrac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1-1)$

$$\frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2(n_2-1)$$

且S_1^2, S_2^2独立，则由F分布的定义知

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

例1：设总体$X\sim N(10,3^2)$, X_1，…，X_n是它的一个样本

$Z=\sum_{i=1}^{6}X_i$ (1)写出Z所服从的分布;(2)求P(Z>11).

解: 1)$\because$ $X\sim N(10,3^2)$，$\therefore$ $Z=\sum_{i=1}^{6}X_i\sim N(60,54)$

$E(Z)=\sum E(X_i)=6\times 10=60$　$D(Z)=\sum D(X_i)=6\times 3^2=54$

2) $\because$ $Z\sim N(60,54)$，

$$\therefore P\{Z>11\}=1-P\{Z\le 11\}=1-\Phi\left(\frac{11-60}{\sqrt{54}}\right)$$

$$=1-\Phi(-6.67)=\Phi(6.67)\approx 1$$

例2：设X_1，…，X_{10}是取自N（0，0.3^2)的样本，

求 $P\{\sum_{i=1}^{10} X_i^2 > 1.44\}$

解：∵ X_1，…，X_{10}～N（0，0.3^2)

$\therefore \frac{X_i}{0.3} \sim$ N（0，1)　　$\sum_{i=1}^{10} \frac{X_i^2}{0.3^2} \sim \chi^2(10)$

$$P\left\{\sum_{i=1}^{10} X_i^2 > 1.44\right\} = P\left\{\sum_{i=1}^{10} \frac{X_i^2}{0.3^2} > \frac{1.44}{0.3^2}\right\}$$

$$= P\left\{\sum_{i=1}^{10} \frac{X_i^2}{0.3^2} > 16\right\} \approx P\left\{\sum_{i=1}^{10} \frac{X_i^2}{0.3^2} > \chi_{0.1}^2(10)\right\} = 0.1$$

例3：设X_1，…，X_n是取自X的样本, 记μ =EX, DX=σ^2，求样本方差S^2的期望。

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

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

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

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

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

$$E(X_i^{\ 2})=D(X_i)+(E(X_i))^2=\sigma^2+\mu^2$$

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

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

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

注: 结论与总体服从的分布无关.

6-1（2001年，数学三）设总体X服从正态分布$N(0,2^2)$**,而** $X_1, X_2, \cdots, X_{15}$**是来自总体X的简单随机样本, 则随机变量**

$$Y=\frac{X_1^2+\cdots+X_{10}^2}{2(X_{11}^2+\cdots+X_{15}^2)}$$

服从_____分布,参数为_______

F $\underline{(10,5).}$

$$\frac{X_1^2+X_2^2+\cdots X_{10}^2}{4}\sim\chi^2(10),\quad \frac{X_{11}^2+\cdots X_{15}^2}{4}\sim\chi^2(5)$$

$$\frac{(X_1^2+\cdots+X_{10}^2)/10}{(X_{11}^2+\cdots+X_{15}^2)/5}=Y\sim F(10,5)$$

6-2（2003年，数学一) 设随机变量 $X \sim t(n)(n>1)$,

$$Y=\frac{1}{X^2}$$ **，则**

(A) $Y \sim \chi^2(n).$ **(B)** $Y \sim \chi^2(n-1).$

(C) $Y \sim F(n,1).$ **(D)** $Y \sim F(1,n).$

[C]