西北工業大學

NORTHWESTERN POLYTECHNICAL UNIVERSITY

概率论与数理统计

徐 爽

西北工业大学理学院概率统计教研室

第五章 数理统计的基本概念与抽样分布

第一节　基本概念

- 一、问题的提出
- 二、总体与个体
- 三、随机样本的定义
- 四、统计量

一、问题的提出

概率论 假设：**研究对象的分布已知**

$$X \sim F(x), p(x), P(x = x_k) \Rightarrow EX, DX, P(A)$$

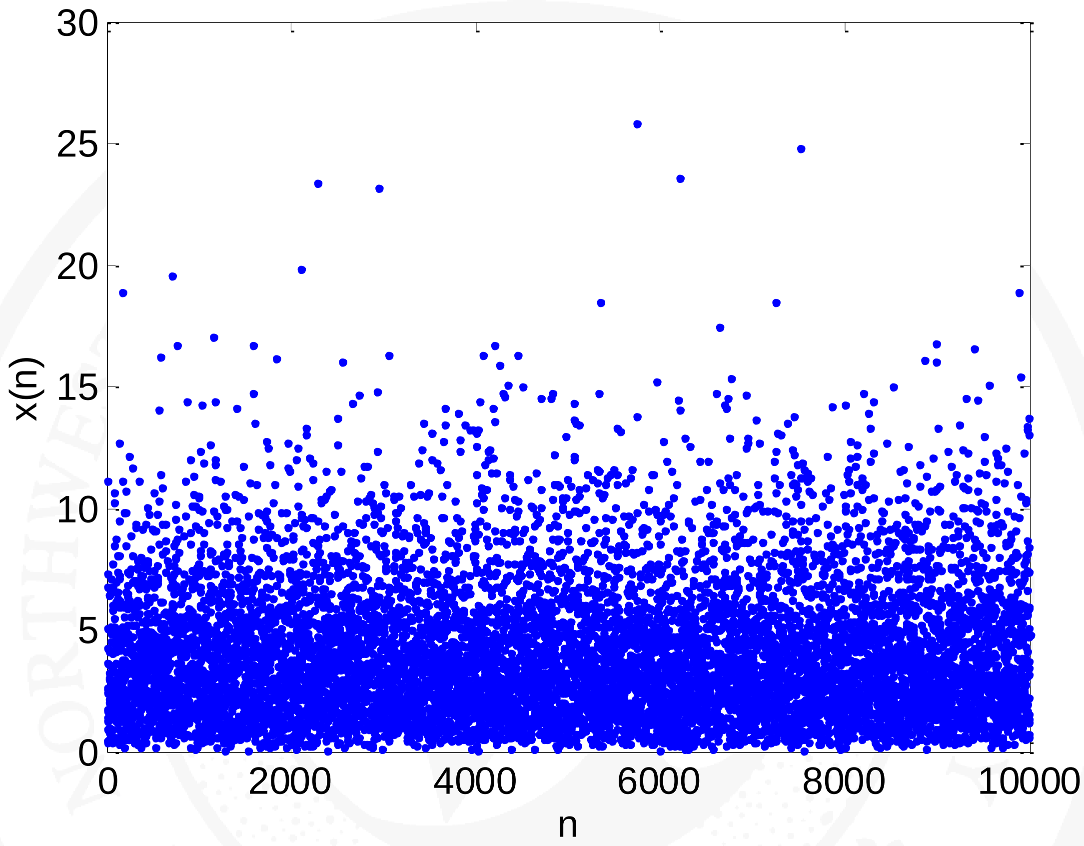
数理统计 实际：**研究对象的分布未知**

需要用已有的**部分信息**去**推断整体**情况。

$$(x_1, x_2, \cdots, x_n) \Rightarrow EX, DX, F(x)$$

什么是统计？

就是瞎子摸象



$$(x_1, x_2, \cdots, x_{10000}) \to EX, DX, F(x,\lambda)?$$

一叶落而知秋天下，一粒米见一世界

二、总体与个体

总体： 在数理统计中，把研究对象的全体称为总体（或母体）.

个体： 总体中每个研究对象称为个体.

在实际中，我们并不关心总体的各个方面，而往往关心它的**某项或几项数量指标**.

例如，在考察我校某届本科生**学习质量**时，该届本科生的**全体成绩**称为**总体**，每一个本科生的成绩称为**个体**.

当我们说到**总体**，就是指数量指标（具有确定概率分布的随机变量）**可能取值**的全体。每一个可能的取值为个体。

总体 ←----→ 随机变量

定义5.1 一个随机变量或者其相应的分布函数 $F(x)$ 称为一个总体.

通常，我们用随机变量 $X, Y, Z \cdots$ 等表示**总体**.

三、随机样本的定义

1. 样本的定义

从总体X中，随机地抽取n个个体：

$$X_1, X_2, \cdots, X_n$$

称为总体X的一个**样本**，记为

$$(X_1, X_2, \cdots, X_n)$$

样本中所包含个体的总数n称为**样本容量**.

注 样本$(X_1, X_2, \cdots, X_n)$是一个n维随机变量.

例 1 为了了解数学专业本科毕业生的月薪情况，调查了某地区100名2013届数学专业的本科生的月薪情况，试问

（1）什么是总体？

（2）什么是样本？

（3）样本容量是多少？

解 （1）总体是该地区2013届数学本科毕业生的月薪；

（2）样本是被调查的100名2013届数学本科毕业生的月薪；

（3）样本容量是100.

2. 样本值

每一次抽取 $X_1, X_2, \cdots, X_n$ 所得到的n个确定的具体数值，记为

$$(x_1, x_2, \cdots, x_n)$$

称为样本

$$(X_1, X_2, \cdots, X_n)$$

的一个**样本值**(观察值).

■ 数理统计的**基本任务**是：根据从总体中抽取的样本，利用样本的信息推断总体的性质。

3. 简单随机样本

若来自总体X的样本$(X_1, X_2, \cdots, X_n)$具有下列两个特征:

(1) 代表性: $X_1, X_2, \cdots, X_n$ 中每一个体与总体X有相同的分布.

(2) 独立性: $X_1, X_2, \cdots, X_n$是相互独立的随机变量.

则称 $(X_1, X_2, \cdots, X_n)$ 为n维简单随机样本.

获得简单随机样本的抽样方法称为**简单随机抽样**.

样本的严格数学定义：

定义5.2 设随机变量X的分布函数为$F(x)$，若$X_1, X_2, \cdots, X_n$是具有同一分布函数$F(x)$，且相互独立的随机变量，则称$X_1, X_2, \cdots, X_n$为来自总体X的容量为n的**简单随机样本**，简称样本.

4. 样本的分布

定理5.1 设$(X_1, X_2, \cdots, X_n)$为来自总体X的样本.

(1)若总体X的分布函数为$F(x)$,则样本

$(X_1, X_2, \cdots, X_n)$的分布函数为$\prod_{i=1}^{n} F(x_i)$

(2)若总体X的分布密度为$p(x)$,则样本

$(X_1, X_2, \cdots, X_n)$的分布密度为$\prod_{i=1}^{n} p(x_i)$

(3)若总体X的分布律为$P(X=x_i)=p(x_i)(i=1,\cdots,n)$

则样本$(X_1,X_2,\cdots,X_n)$的分布律为$\prod_{i=1}^{n}p(x_i)$.

样本的分布：

***n*维独立同分布随机变量的联合分布**

四、统计量

$$(x_1, x_2, \cdots, x_{10000}) \to F(x, \lambda)$$

由样本推断总体情况,需要对样本值进行“加工”,这就需要构造一些**样本的函数**,它把样本中所含的信息集中起来.

1. 统计量

定义5.3 设$(X_1 \cdots X_2, \cdots, X_n)$是来自总体$X$的一个样本，$f(X_1, X_2, \cdots, X_n)$是$X_1, X_2, \cdots, X_n$的函数，若$f$中不含任何关于总体$X$的未知参数，则称$f(X_1, X_2, \cdots, X_n)$是一个统计量.

设$(x_1, x_2, \cdots, x_n)$是样本$(X_1, X_2, \cdots, X_n)$的观察值

则称$f(x_1, x_2, \cdots, x_n)$是统计量 $f(X_1, X_2, \cdots, X_n)$的观察值

注 1°统计量$Y = f(X_1, X_2, \cdots, X_n)$是随机变量;

2° 统计量用于统计推断，故不应含任何关于总体X的未知参数;

3° 统计量是样本的函数，它是一个**随机变量**，统计量的分布称为**抽样分布**.

2. 几个常用统计量

$f(X_1, X_2, \cdots, X_n)$

(1) 样本矩

设 $X_1, X_2, \cdots, X_n$ 是来自总体的一个样本, $x_1, x_2, \cdots, x_n$ 是这一样本的观察值.

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

它反映了总体均值的信息

其观察值 $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

3.9735， 3.9820， 3.9735…

可用于推断：$E(X)$.

2) 样本方差

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

它反映了总体方差的信息

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

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

其观察值

$$s_n^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2$$

7.8544， 7.7280, 7.8440

可用于推断：$D(X)$.

3) 样本标准差

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

其观察值

$$s_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}.$$

4) 修正样本方差

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

其观察值

$$s_n^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right).$$

样本方差与修正样本方差的**关系:**

$$S_n^2 = \frac{n-1}{n}S_n^{*2}. \quad \le S_n^{*2}$$

注 1° 当n较大时， S_n^{*2}与S_n^2 差别微小；

2° 当n较小时,S_n^{*2}比S_n^2 有更好的统计性质.

5) 样本 k 阶(原点)矩

$$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k,\ k=1,2,\cdots;$$ 特例：$A_1=\overline{X}$

其观察值 $$a_k=\frac{1}{n}\sum_{i=1}^{n}x_i^k,\ k=1,2,\cdots.$$

6)样本 k 阶中心矩

特例：$B_2=S_n^2$

$$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k,\ k=2,3,\cdots;$$

其观察值 $$b_k=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^k,\ k=2,3,\cdots.$$

样本矩具有下列性质:

性质5.1 设总体X的期望$EX=\mu$,方差$DX=\sigma^2$,

$(X_1,X_2,\cdots,X_n)$为来自总体X的样本，则有

$$(1)\quad E(\bar{X})=E(X)=\mu$$

$$(2)\quad D(\bar{X})=\frac{1}{n}D(X)=\sigma^2;$$

$$(3)\quad E(S_n^2)=\frac{n-1}{n}D(X)=\frac{n-1}{n}\sigma^2;$$

$$(4)\quad E(S_n^{*2})=D(X)=\sigma^2.$$

证 (1) $E(\overline{X}) = \mu$

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

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

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

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

$$(3)\ E(S_n^2)=\frac{n-1}{n}\sigma^2$$

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

$$=\frac{1}{n}\sum_{i=1}^{n}\{D(X_i)+[E(X_i)]^2\}-\{D(\bar{X})+[E(\bar{X})]^2\}=$$

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

$$(4)\ E(S_n^{*2})=E(\frac{n}{n-1}S_n^2)=\frac{n}{n-1}E(S_n^2)=\sigma^2$$

性质5.2 若总体X的k阶矩$E(X^k)=a_k$存在，

则当$n\to\infty$时，$A_k=\dfrac{1}{n}\sum_{i=1}^{n}X_i^k\xrightarrow{p}a_k, k=1,2,\cdots$

证 因为 $X_1,X_2,\cdots,X_n$ 独立且与 X 同分布,

所以 $X_1^k,X_2^k,\cdots,X_n^k$ 独立且与 X^k 同分布,

故有 $E(X_1^k)=E(X_2^k)=\cdots=E(X_n^k)=a_k.$

再根据第四章辛钦大数定理，即

若$r.v$ $X_1\cdots X_n$独立同分布，且$EX_k=\mu$,则

$\forall\varepsilon>0$ 有$\dfrac{1}{n}\sum_{i=1}^{n}X_i\xrightarrow{P}EX_i$

由上述定理可得

$$\frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{p} a_k\ , \quad k = 1, 2, \cdots;$$

$$\Rightarrow \overline{X} \xrightarrow{p} EX;$$

由第四章关于依概率收敛的序列的性质知

$$g(A_1, A_2, \cdots, A_k) \xrightarrow{P} g(a_1, a_2, \cdots, a_k),$$

其中g是连续函数.

$$\Rightarrow S_n^2 = A_2 - A_1^2 \xrightarrow{P} DX = a_2 - a_1^2;$$

注 性质5.2是下一章矩估计法的理论根据.

(2) 次序统计量

设$(X_1, X_2, \cdots, X_n)$是从总体X中抽取的一个样本, $(x_1, x_2, \cdots, x_n)$是其一个观测值,将观测值按由小到大的次序重新排列为

$$x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$$

当$(X_1, X_2, \cdots, X_n)$取值为$(x_1, x_2, \cdots x_n)$时,定义

$X_{(k)}$取值为$x_{(k)}(k = 1,2,\cdots n)$,由此得到

$$(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$$

称为样本 $(X_1, X_2, \cdots, X_n)$的**次序统计量.**

对应的$(x_{(1)}, x_{(2)}, \cdots x_{(n)})$称为其观测值.

$X_{(k)}$： 样本$(X_1, X_2, \cdots, X_n)$的第k个次序统计量.

特别地， $X_{(1)} = \min\limits_{1 \le i \le n} X_i$ 称为最小次序统计量.

$X_{(n)} = \max\limits_{1 \le i \le n} X_i$ 称为最大次序统计量.

注 由于每个$X_{(k)}$都是样本$(X_1, X_2, \cdots, X_n)$的函数，

所以，$X_{(1)}, X_{(2)}, \cdots, X_{(n)}$也是随机变量，但它们

一般不相互独立.

$$(X_1, X_2, X_3, X_4, X_5) \qquad X_k \sim N(2,3)$$

第1次抽样	-1.6225	4.1517	6.8907	3.4667	5.1041
第2次抽样	2.9756	-0.2648	6.1109	-3.1345	1.6933
第3次抽样	-1.4412	-1.2066	-0.4285	-6.8329	6.3151

$$(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$$

-1.6225	3.4667	4.1517	5.1041	6.8907
-3.1345	-0.2648	1.6933	2.9756	6.1109
-6.8329	-1.4412	-1.2066	-0.4285	6.3151

定理5.2 设总体X的分布密度为$p(x)$(或分布函数为$F(x)$),$(X_{(1)},X_{(2)},\cdots,X_{(n)})$为总体$X$的样本$(X_1,X_2,\cdots,X_n)$的次序统计量.则有

(1) 最大次序统计量$X_{(n)}$的分布密度为

$$p_{X_{(n)}}(x)=n[F(x)]^{n-1}p(x).$$

(2) 最小次序统计量$X_{(1)}$的分布密度为

$$p_{X_{(1)}}(x)=n[1-F(x)]^{n-1}p(x).$$

证明思路：极值分布

证 (1) $F_{X_{(n)}}(x) = P\{X_{(n)} \le x\}$

$$= P\{\max_{1\le i\le n} X_i \le x\}$$

$$= P\{X_1 \le x, X_2 \le x, \cdots, X_n \le x\}$$

$$= P\{X_1 \le x\} \cdot P\{X_2 \le x\} \cdot \cdots \cdot P\{X_n \le x\}$$

$$= F^n(x)$$

$$\therefore \quad p_{X_{(n)}}(x) = \frac{\mathrm{d} F_{X_{(n)}}(x)}{\mathrm{d} x} = nF^{n-1}(x) \cdot p(x)$$

(3) 经验分布函数

定义5.5 设 $X_1, X_2, \cdots, X_n$ 是总体 X 的一个样本, $(X_{(1)}, X_{(2)}, \cdots, X_{(n)})$ 为总体 X 的样本 $(X_1, X_2, \cdots, X_n)$ 的次序统计量.

$(x_{(1)}, x_{(2)}, \cdots x_{(n)})$ 为其观测值, 设 x 是任一实数, 称函数

$$F_n(x) = \begin{cases} 0, & x < x_{(1)}, \\ \dfrac{k}{n}, & x_{(k)} \le x < x_{(k+1)}, \\ 1, & x \ge x_{(n)}. \end{cases}$$

为总体X的**经验分布函数**，即对于任何实数 x 经验分布函数 $F_n(x)$ 为样本值中不超过x的个数再除以n，亦即

$$F_n(x)=\frac{\mu_n(x)}{n}$$

其中$\mu_n(x)\,(-\infty<x<+\infty)$表示 $x_1,x_2,\cdots,x_n$ 中不超过于 x 的个数.

注 1° $\mu_n(x)$ 为样本中不超过x的样本的最大个数，即在n次重复独立试验中，事件

$A=\{X \le x\}$ 发生的次数. $P(A)=F(x)$

$(\because\ x_{(1)} \le x_{(2)} \le \cdots \le x_{(k)} \le x$, 有$\mu_n(x)$个样品的取值$\le x)$

2° $F_n(x)=\dfrac{\mu_n(x)}{n}$ 为事件$\{X \le x\}$发生的频率.

事实上，令 $\mu_n(x)=\sum_{i=1}^{n} I_i$，其中

$$I_i=\begin{cases}1, & \{X_i \le x\}\text{发生}\\ 0, & \{X_i \le x\}\text{不发生}\end{cases} \sim B(1,F(x)),\ \text{则}\ \mu_n(x) \sim B(n,F(x))$$

性质

(1)对于给定的一组样本值$(x_1, x_2, \cdots, x_n)$，$F_n(x)$满足分布函数的特征：$0 \le F_n(x) \le 1, F_n(-\infty) = 0, F_n(+\infty) = 1,$ 单调非降右连续，是一个分布函数.

(2)由于$F_n(x)$是样本的函数,故$F_n(x)$是随机变量.

可以证明$nF_n(x) = \sum^{n} I_i \sim B(n, F(x))$,所以

$$E[F_n(x)] = F(x), \quad D[F_n(x)] = \frac{F(x)[1-F(x)]}{n}$$

(3)$F_n(x)$依概率收敛于$F(x)$. 即

$$\lim_{n\to\infty} P\{|F_n(x) - F(x)| < \varepsilon\} = 1 \qquad (\forall \varepsilon > 0)$$

例10 设从总体 X 中取得一个容量为5的样本，样本观测值为 -2，-1，2.5，3.1，3.7，试求此样本经验分布分布函数 $F(x)$.

解 由经验分布函数的定义知

$$F(x)=\begin{cases} 0 & x<-2 \\ 1/5 & -2\le x<-1 \\ 2/5 & -1\le x<2.5 \\ 3/5 & 2.5\le x<3.1 \\ 4/5 & 3.1\le x<3.7 \\ 1 & 3.7\le x \end{cases}$$

内容小结

基本概念： 总体 X (随机变量)

个体 $X_1, X_2, \cdots, X_n$ （随机变量）

（简单随机）样本 $(X_1, X_2, \cdots, X_n)$ （n维随机向量）

样本值 $(x_1, x_2, \cdots, x_n)$ （常量）

总体、样本、样本值的相互关系：

X 总体(理论分布)

估计

$(X_1, X_2, \cdots, X_n)$ 样本 → 样本值 $(x_1, x_2, \cdots, x_n)$

统计量: $f(X_1, X_2, \cdots, X_n)$ 随机变量

观测值: $f(x_1, x_2, \cdots, x_n)$ 常量

1、样本矩

样本均值 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} \mu = EX$

(1) $E(\bar{X}) = \mu$ (2) $D(\bar{X}) = \frac{1}{n}\sigma^2$;

样本方差 $S_n^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \xrightarrow{p} \sigma^2 = DX$

(3) $E(S_n^2) = \frac{n-1}{n}\sigma^2$; (4) $E(S_n^{*2}) = \sigma^2$.

样本矩 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{p} \mu_k = E(X^k)$

2、次序统计量:

(1) 最大次序统计量$X_{(n)}$的分布密度为

$$p_{X_{(n)}}(x)=n[F(x)]^{n-1}p(x).$$

(2) 最小次序统计量$X_{(1)}$的分布密度为

$$p_{X_{(1)}}(x)=n[1-F(x)]^{n-1}p(x).$$

3、经验分布函数:

$$F_n(x)=\frac{\mu_n(x)}{n}\xrightarrow{p,a.e.}F(x)$$

$$nF_n(x)=\mu_n(x)=\sum_{i=1}^{n}I_i\sim B(n,F(x))$$

$$E[F_n(x)]=F(x),\quad D[F_n(x)]=\frac{F(x)[1-F(x)]}{n}$$

5-1 基本概念

Thank You !

西北工業大學
NORTHWESTERN POLYTECHNICAL UNIVERSITY

概率论与数理统计

徐 爽

西北工业大学理学院概率统计教研室

第五章 数理统计的基本概念与抽样分布

第一节 基本概念

第二节 常用统计分布

第三节 抽样分布

第二节 常用统计分布

- 一、常见分布
- 二、概率分布的分位数

一、常见分布

1. χ^2 分布

(1) 定义

定义5.6 设$X_1, X_2, \cdots, X_n$是来自总体$N(0,1)$的样本，则称统计量$\chi_n^2 = X_1^2 + X_2^2 + \cdots + X_n^2$服从自由度为$n$的$\chi^2$分布. $\sim \chi^2(n)$

自由度n:

指 $\chi_n^2 = X_1^2 + X_2^2 + \cdots + X_n^2$ 中右端包含独立变量的个数.

(2) χ_n^2分布的概率分布

χ_n^2分布的概率密度:

$$p(x)=\begin{cases}\dfrac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}x^{\frac{n}{2}-1}e^{-\frac{x}{2}} & x>0\\ 0 & \text{其它}\end{cases}$$

证

$$\Gamma(\alpha,\beta)\text{分布}: p(x)=\begin{cases}\dfrac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}, & x>0\\ 0, & x\le 0\end{cases}$$

$$\therefore \Gamma(\frac{n}{2},\frac{1}{2})=\chi^2(n)$$

其中：$\Gamma(\alpha)=\int_0^{+\infty}x^{\alpha-1}e^{-x}dx \quad (\alpha>0)$

$$\Gamma(\alpha+1)=\alpha\Gamma(\alpha) \Rightarrow \Gamma(n+1)=n!$$

$$\Gamma(1)=1,\Gamma(\frac{1}{2})=\sqrt{\pi}$$

特别的$X \sim N(0,1)$, 则$X^2 \sim \chi^2(1)=\Gamma\left(\frac{1}{2},\frac{1}{2}\right)$.

$$\Rightarrow \quad \chi^2(n)=\Gamma(\frac{n}{2},\frac{1}{2}) \qquad \Gamma(1,\beta)=Exp(\beta)$$

$$\therefore \chi^2(2)=\Gamma(1,\frac{1}{2})=Exp(\frac{1}{2})$$

(3) χ^2 分布的性质

性质5.3 (χ^2 分布的可加性)

设 $Y_1 \sim \chi^2(n_1)$, $Y_2 \sim \chi^2(n_2)$, 并且 Y_1, Y_2 独立, 则 $Y_1 + Y_2 \sim \chi^2(n_1 + n_2)$.

(此性质可以推广到多个随机变量的情形)

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

性质5.4 （χ^2分布的数学期望和方差）

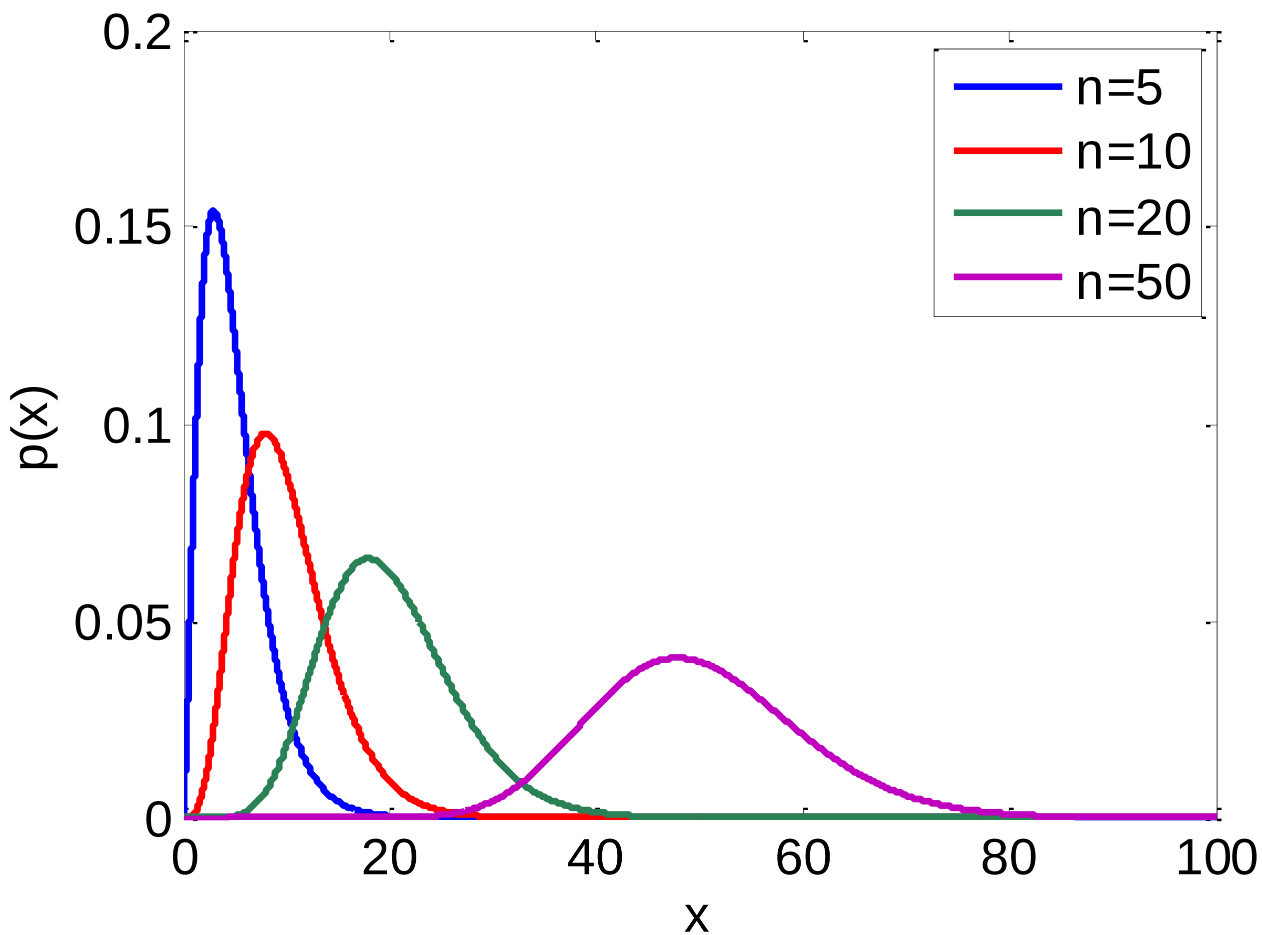
若 $\chi_n^2 \sim \chi^2(n)$, 则 $E(\chi_n^2) = n, \quad D(\chi_n^2) = 2n.$

性质5.5 设$\chi_n^2 \sim \chi^2(n)$,则对任意x,有

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

即$\chi_n^2 \sim AN(n, 2n)$ (开方分布的极限分布为正态分布)

$$X \sim \chi^2(n)$$

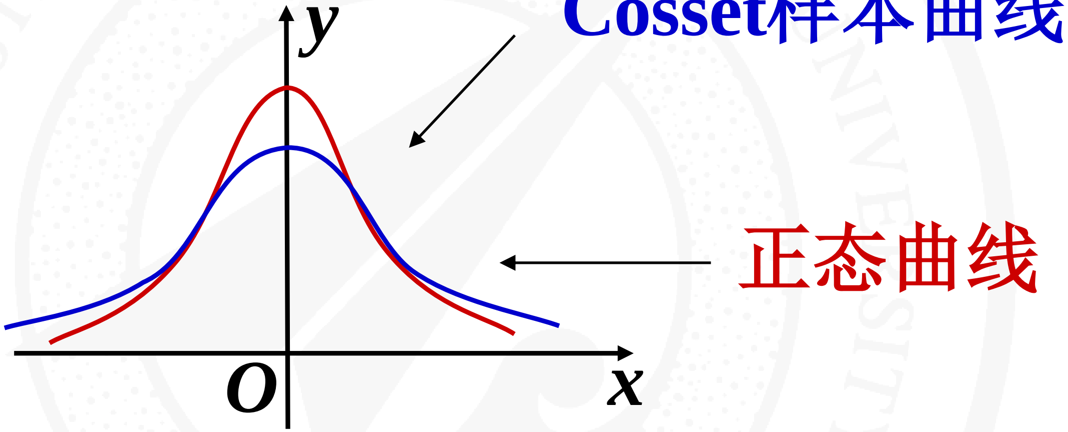


2. *t* 分布

历史上，正态分布由于其广泛的应用背景和良好的性质，曾一度被看作是“**万能分布**”，我们知道在总体均值和方差已知情况下，**样本均值的分布将随样本量增大而接近正态分布，即** $\overline{X} \sim AN(\mu, \frac{1}{n}\delta^2).$

在这样的背景下，十九世纪初英国一位年轻的酿酒化学技师**Cosset. WS**, 他在酒厂从事试验数据分析工作，对数据误差有着大量感性的认识，

但是Cosset在实验中遇到的样本容量仅有5~6个，在其中他发现实际数据的分布情况与正态分布有着较大的差异.



于是Cosset怀疑存在一个不属于正态的其他分布，通过学习终于得到了新的密度曲线，并在1908年以“Student”笔名发表了此项结果，后人称此分布为“t 分布”或“学生氏”分布.

(1) 定义

定义5.7 设 $X \sim N(0,1),\ Y \sim \chi^2(n)$, 且 X, Y 独立, 则称随机变量

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

服从自由度为 n 的 t 分布, 记为 $T \sim t(n)$.

t 分布又称**学生氏**

(**Student**)**分布**.

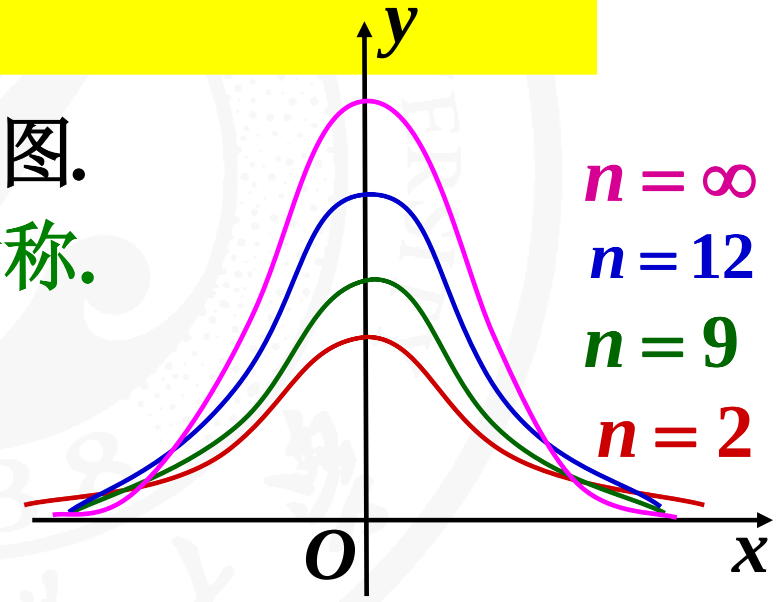
(2) $t(n)$ 分布的概率密度函数为

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

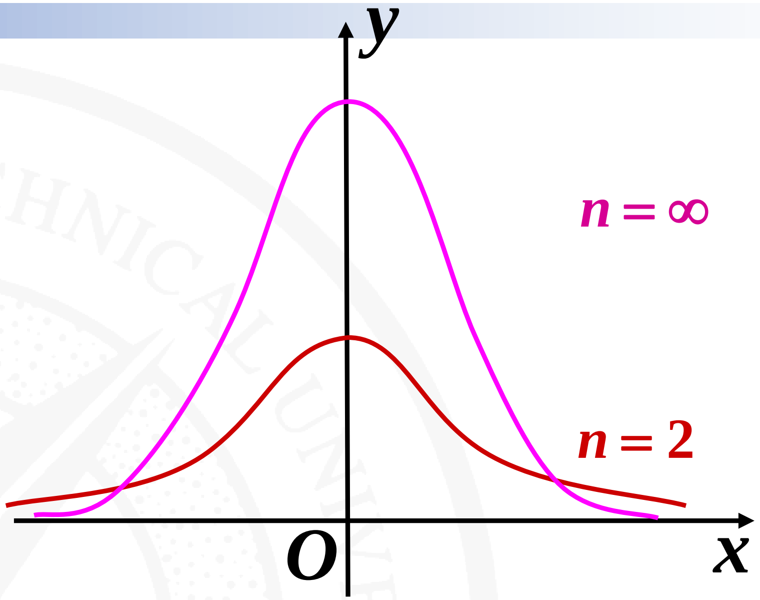
t分布的概率密度曲线如图.

显然图形是关于 $t=0$ 对称.

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.



(3) t分布的性质



性质5.6 若$T \sim t(n)$，则

$$E(T)=0,$$

$$D(T)=\frac{n}{n-2} \qquad (n>2).$$

性质5.7 若$T \sim t(n)$，则

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

即$T \sim AN(0,1)$分布

但对于较小的n，t分布与$N(0,1)$分布相差很大，具有厚尾性。

3. F分布

(1) 定义

定义5.8 设$X \sim \chi^2(n_1)$,$Y \sim \chi^2(n_2)$,且X,Y相互独立,

则称随机变量

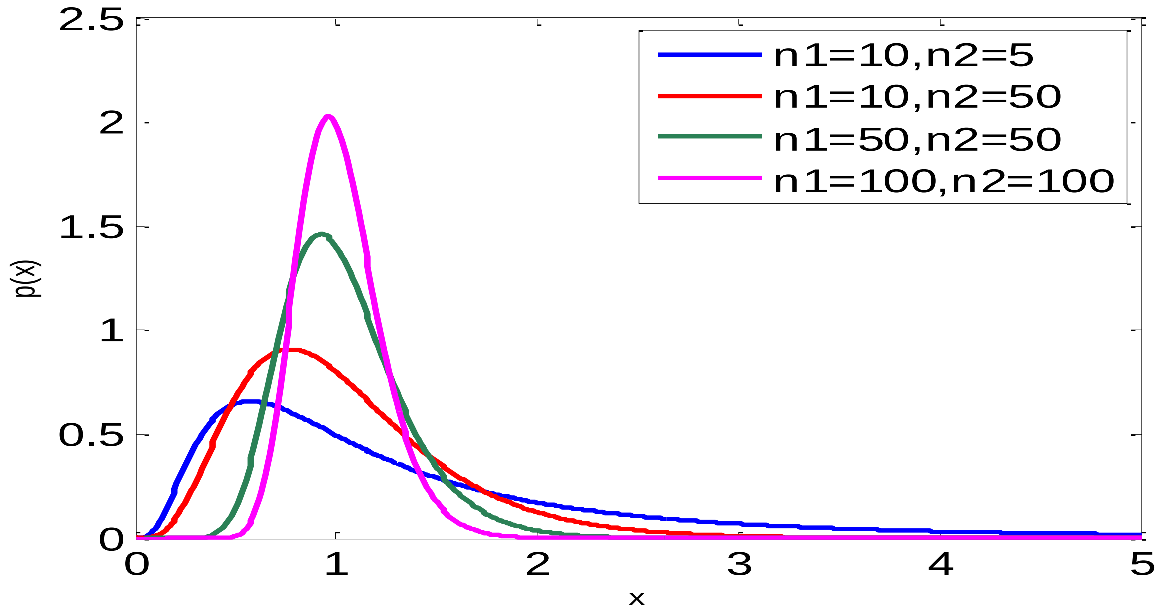
$$F=\frac{X/n_1}{Y/n_2}$$

服从自由度为(n_1,n_2)的F分布，记为

$$F \sim F(n_1,n_2).$$

(2) $F(n_1, n_2)$分布的概率密度为

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



(3) F分布有以下性质

1) $$E(F)=\frac{n_2}{n_2-2},\quad (n_2>2),$$

$$D(F)=\frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)},\quad (n_2>4)$$

2) **性质5.8** 若$F\sim F(n_1,n_2)$, 则$\dfrac{1}{F}\sim F(n_2,n_1)$.

3) 设$F\sim F(n_1,n_2)$,则当$n_2>4$时,对任意x有

$$\lim_{n_1\to\infty}P\left\{\frac{F-E(F)}{\sqrt{D(F)}}\le x\right\}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt$$

这说明F分布极限分布也是正态分布.

$T \sim t(n)$,则$T^2 \sim F(1,n)$.

证 因为$T \sim t(n)$, 由定义5.7有

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

其中$X \sim N(0,1), Y \sim \chi^2(n)$,且$X, Y$独立,

从而 $X^2 \sim \chi^2(1), Y \sim \chi^2(n)$, 且$X^2$与$Y$独立,

$\therefore$ 由定义5.8,有 $T^2 = \dfrac{X^2}{Y/n} \sim F(1,n)$.

例1 设$X \sim N(0,4), Y \sim \chi^2(2)$，且$X, Y$相互独立，试求解 $\dfrac{X^2}{4}+Y$ 的概率分布.

解 因为$X \sim N(0,4)$且 X, Y 相互独立，所以

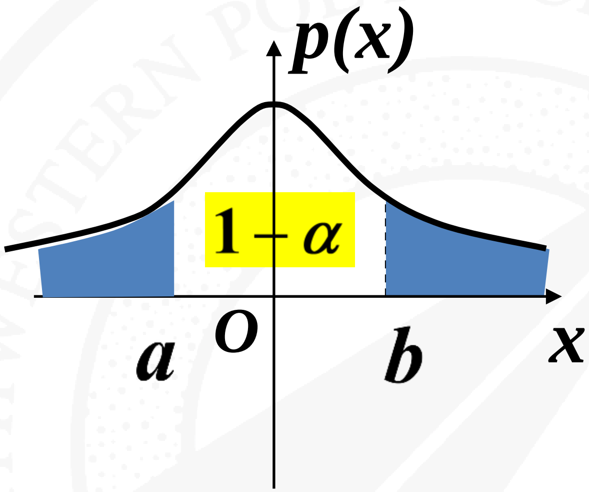
$$\frac{X}{2} \sim N(0,1)$$

且 $\dfrac{X^2}{4}$与Y相互独立

又因为 $\dfrac{X^2}{4} \sim \chi^2(1)$，由开方分布的可加性得

得 $$\frac{X^2}{4}+Y \sim \chi^2(3).$$

二、概率分布的分位数



已知X的分布 $\Leftrightarrow P(a<X\le b)=1-\alpha$

(1)给定区间范围a,b,求概率 $P(a<X\le b)=?$

(2)给定概率α求区间范围$[a,b]$

1. 定义

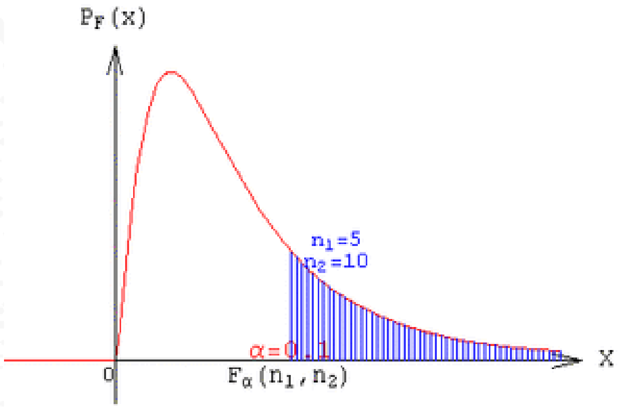
定义5.9 对于总体X和给定的$\alpha(0<\alpha<1)$,

若存在x_α,使

$$P\{X > x_\alpha\} = \alpha$$

则称x_α为X的分布的上侧α分位数.

$$F(x_\alpha) = P\{X \le x_\alpha\} = 1-\alpha$$



2. 常用分布的上侧分位数记号

分布	$N(0,1)$	$\chi^2(n)$	$t(n)$	$F(n_1,n_2)$
记号	u_α	$\chi^2_\alpha(n)$	$t_\alpha(n)$	$F_\alpha(n_1,n_2)$

3. 查表法

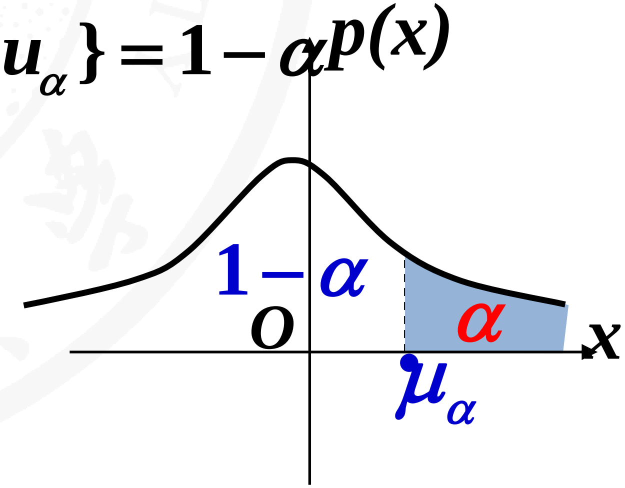
(1) 若X的分布密度关于y轴对称，则

1) 正态分布的上侧分位数u_α:

设$X \sim N(0,1)$,则其上侧分位数u_α满足 $P\{X > u_\alpha\} = \alpha$

$$\Phi(u_\alpha) = P\{X \le u_\alpha\} = 1 - P\{X > u_\alpha\} = 1 - \alpha$$

即 $\Phi(u_\alpha) = 1 - \alpha$



给定α,由附表2可查得u_α的值.

$$\boxed{\Phi(u_\alpha) = P(X \le u_\alpha) = 1-\alpha}$$

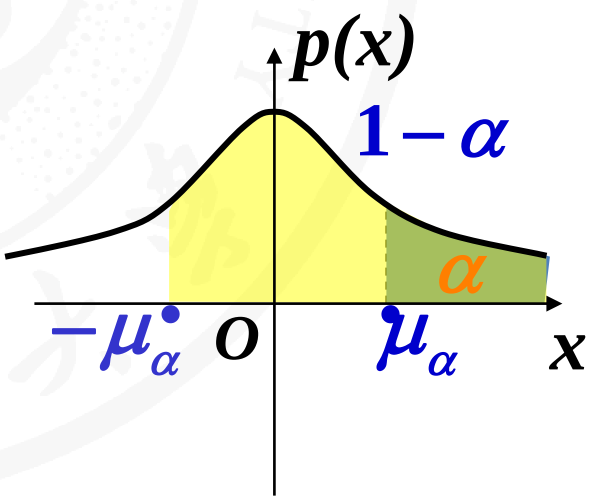
$u_{0.05} = 1.645$, $\boxed{\Phi(1.645) = 0.95}$ $(\alpha = 0.05)$

$u_{0.025} = 1.96$, $\boxed{\Phi(1.96) = 0.975}$ $(\alpha = 0.025)$

根据正态分布的对称性知 $u_{1-\alpha} = -u_\alpha$.

$\therefore$ 当$\alpha > 0.5$无法查表

$$u_\alpha = -u_{1-\alpha}.$$



例: $u_{0.95} = -u_{0.05} = -1.645.$

2) t分布的上侧分位数$t_\alpha(n)$:

设$T\sim t(n)$,则其上侧分位数t_α满足

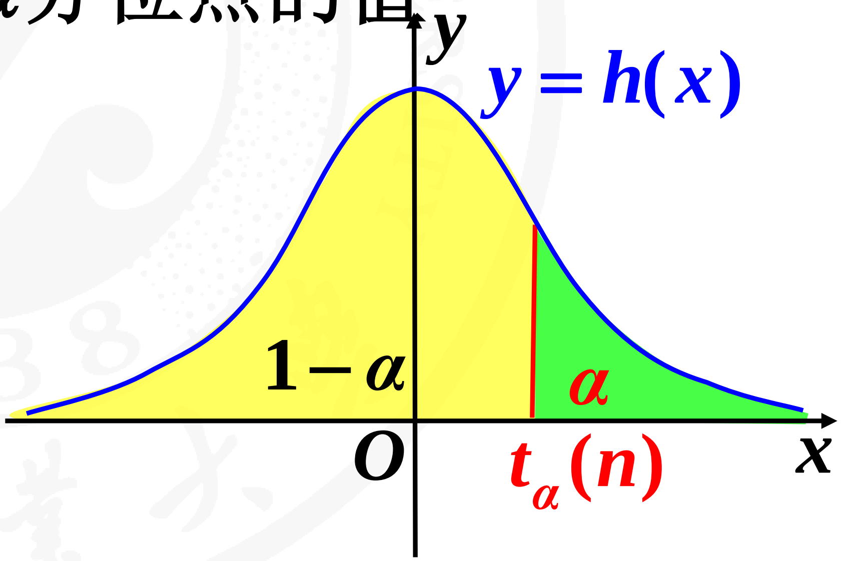
$$P\{T>t_\alpha(n)\}=\alpha$$

可以通过查表求得上α分位点的值

$(\alpha\le 0.25, n\le 45)$.

$t_{0.05}(10)=1.8125,$

$t_{0.025}(15)=2.1315.$



当$\alpha > 0.25, n \le 45,$ 由分布的对称性知

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

$$t_{0.95}(10) = -t_{0.05}(10) \quad = -1.8125,$$

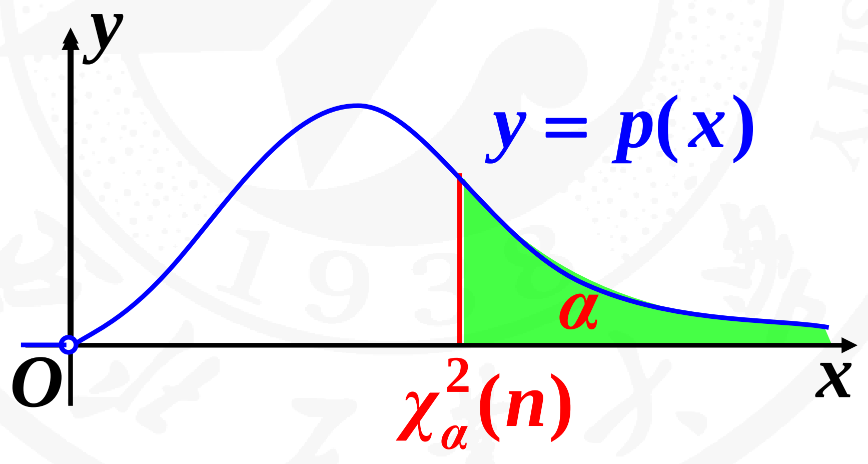
当$n > 45$时，$t_{\alpha}(n) \approx u_{\alpha}.$

$$t_{0.05}(100) \approx \mu_{0.05} \quad = 1.645$$

(2) X的分布密度无对称性的情形

1)设$\chi^2 \sim \chi^2(n)$,则其上侧分位数$\chi_\alpha^2(n)$满足

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

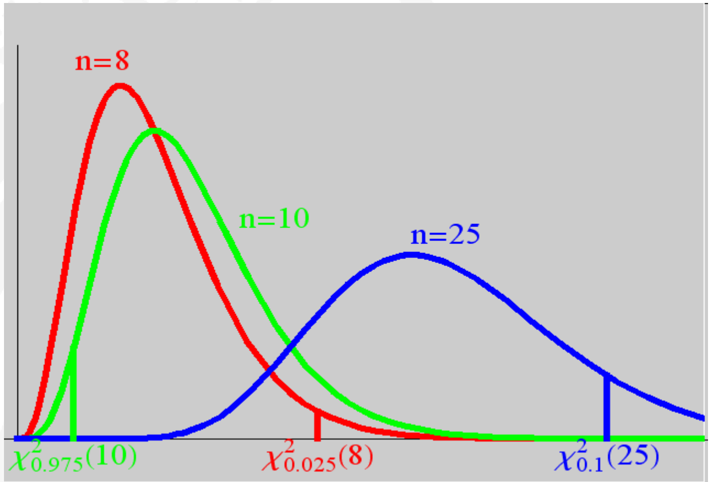


当$n \le 60$时，可查表4(表4只详列到 n=60 为止).

$$\chi^2_{0.025}(8) = 17.5,$$

$$\chi^2_{0.975}(10) = 3.25,$$

$$\chi^2_{0.1}(25) = 34.382.$$



2) F分布的上侧分位数$F_\alpha(n_1, n_2)$:

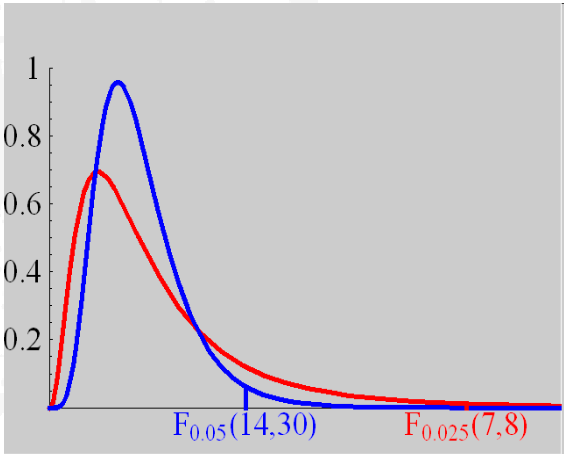
对于 $\alpha = 0.01,\ 0.025,\ 0.05,\ 0.1$ 等， 可直接查表5 ~ 8.

$F_{0.05}(14,30) = 2.04.$ $F_{0.025}(7,8) = 4.53,$

当α为其它值，可利用关系式

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

由F_α求得$F_{1-\alpha}$.



如： $F_{0.95}(12,9) = \dfrac{1}{F_{0.05}(9,12)} = \dfrac{1}{2.8} = 0.357$.

内容小结

1.三大抽样分布:

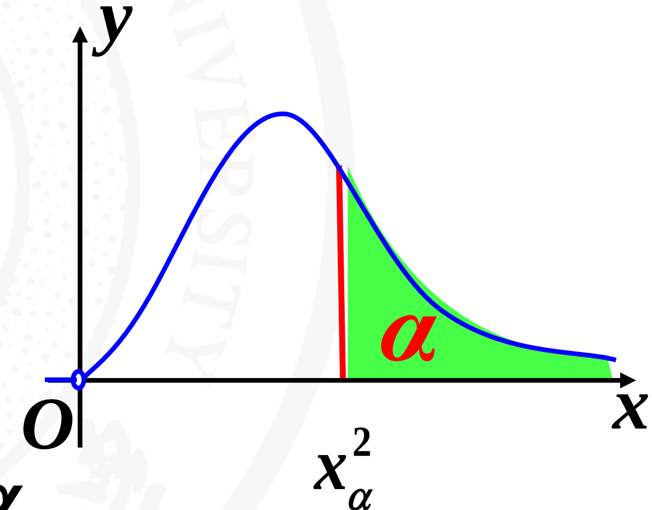
χ^2 分布, t 分布, F 分布

的定义,性质.

2.概率分布的分位数概念. x_α

$$P\{X > x_\alpha\} = \alpha$$

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



$$u_{1-\alpha} = -u_\alpha; \quad t_{1-\alpha}(n) = -t_\alpha(n) \quad F_\alpha(n_1, n_2) = \frac{1}{F_{1-\alpha}(n_2, n_1)}.$$

(1) χ^2 分布 $X_1, X_2, \cdots, X_n \sim N(0,1)$且$X_i$相互独立，

$$\chi_n^2 = X_1^2 + X_2^2 + \cdots + X_n^2 \sim \chi^2(n)$$

$$\Rightarrow E(\chi_n^2) = n, \quad D(\chi_n^2) = 2n$$

(2) t 分布 $X \sim N(0,1)$, $Y \sim \chi^2(n)$, 且 X, Y独立,

则$T = \dfrac{X}{\sqrt{Y/n}} \sim t(n)$ $\Rightarrow E(T) = 0,\ D(T) = \dfrac{n}{n-2}\ (n > 2).$

(3) F 分布

设$X \sim \chi^2(n_1)$, $Y \sim \chi^2(n_2)$, 且X, Y相互独立，

则$F = \dfrac{X/n_1}{Y/n_2} \sim F(n_1, n_2)$。

若$T \sim t(n)$, 则 $T^2 \sim F(1, n)$

分布	统计量	上侧分位数	性质
$N(0,1)$	U	μ_α	$\mu_{1-\alpha}=-\mu_\alpha$, $\Phi(\mu_\alpha)=1-\alpha$
$\chi^2(n)$	$\chi_n^2=\sum_{i=1}^{n}U_i^2$	$\chi_\alpha^2(n)$	$\chi_\alpha^2(n)\approx n+\sqrt{2n}\mu_\alpha$ $(n>60)$
$t(n)$	$T=\dfrac{U}{\sqrt{\chi_n^2/n}}$	$t_\alpha(n)$	$t_{1-\alpha}(n)=-t_\alpha(n)$, $t_\alpha(n)\approx\mu_\alpha(n>45)$
$F(n_1,n_2)$	$F=\dfrac{\chi_{n_1}^2/n_1}{\chi_{n_2}^2/n_2}$	$F_\alpha(n_1,n_2)$	$F_\alpha(n_1,n_2)=\dfrac{1}{F_{1-\alpha}(n_2,n_1)}$

5-2 常用统计分布

Thank You !

西北工業大學

NORTHWESTERN POLYTECHNICAL UNIVERSITY

概率论与数理统计

徐 爽

西北工业大学理学院概率统计教研室

第五章 数理统计的基本概念与抽样分布

第一节 基本概念

第二节 常用统计分布

第三节 抽样分布

第三节　抽样分布

- 一、问题的提出
- 二、抽样分布定理

一、问题的提出

总体 X： 随机变量 $\sim E(X), D(X), F(x,\lambda)$

例如：成绩、温度、时间、质量、长度

估计

样本 $(X_1, X_2, \cdots, X_n)$： n维随机向量

例如：（第1个学生成绩，第2个学生成绩…第n个学生成绩）

（第1天温度，第2天温度…第n天温度）

样本观测值：$(86, 93, \cdots, 65, 72)$

统计量 $f(X_1, X_2, \cdots, X_n)$： 随机变量

抽样分布：统计量的分布 $f(X_1, X_2, \cdots, X_n) \sim$ 分布？

抽样分布 $\begin{cases} \text{精确抽样分布} & \text{(小样本问题中使用)} \\ \text{渐近分布} & \text{(大样本问题中使用)} \end{cases}$

例如： $\overline{X}$ 平均成绩、平均温度…

S_n^2 成绩的方差、温度的方差…

这一节, 我们来讨论**正态总体**的抽样分布.

二、抽样分布定理

1. 样本来自单个正态总体

定理5.3 设样本$(X_1, X_2, \cdots, X_n)$是来自总体X， 而

$$X \sim N(\mu, \sigma^2),$$

则 (1) 样本均值

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

$$\because E(\bar{X}) = E(X)$$

$$D(\bar{X}) = \frac{D(X)}{n}$$

或

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

标准化样本均值

$$X \sim N(\mu,\sigma^2), \quad \Rightarrow X_i \sim N(\mu,\sigma^2), \quad \overline{X} \sim N(\mu,\sigma^2/n),$$

$$\boxed{令 Y_i = X_i - \mu} \sim N(0,\sigma^2)$$

$$\boxed{令 \overline{Y} = \overline{X} - \mu} \sim N(0,\sigma^2/n)$$

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

$$= \frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \overline{Y}^2$$

$$\boxed{令 P_i = \frac{Y_i}{\sigma}} \sim N(0,1)$$

$$= \frac{\sigma^2}{n}\sum_{i=1}^{n}\left(\frac{Y_i}{\sigma}\right)^2 - \frac{\sigma^2}{n}\left(\frac{\overline{Y}}{\sigma/\sqrt{n}}\right)^2$$

$$\boxed{令 Q = \frac{\overline{Y}}{\sigma/\sqrt{n}}} \sim N(0,1)$$

$$\therefore \frac{S_n^2}{\sigma^2/n} = \sum_{i=1}^{n} P_i^2 - Q^2 \quad \sim \chi(n-1)$$

(2) $V=\dfrac{nS_n^2}{\sigma^2}=\dfrac{(n-1)S_n^{*2}}{\sigma^2}$

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

其中S_n^2是样本方差.

注 1° 减少一个自由度的原因：

$$\{\frac{X_i-\overline{X}}{\sigma}\}(i=1,2\cdots,n)\text{不相互独立.}$$

事实上，它们受到一个条件的约束：

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

$$\therefore \sum_{i=1}^{n} \frac{X_i - \overline{X}}{\sigma} = \frac{1}{\sigma}\left(\sum_{i=1}^{n} X_i - n\overline{X}\right) = \frac{1}{\sigma} \cdot 0 = 0.$$

(3) $\overline{X}$ 与 S_n^2 独立.

注　2° 若X不服从正态分布，由中心极限定理知，

当$n >> 1$ (一般$n \geq 30$)时，

$$U = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \overset{\text{近似}}{\sim} AN(0, 1),$$

其中　$\mu = E(X)$，$\sigma^2 = D(X)$.

3° 在实际问题中，总体方差σ^2常常是未知的，若将标准样本均值U中的σ用S_n^*代替，则有如下推论：

（4） 设$(X_1, X_2, \cdots, X_n)$是总体$N(\mu,\sigma^2)$的样本，$\bar{X}, S_n^{*2}$ 分别是样本均值和修正样本方差，则有

$$T=\frac{\bar{X}-\mu}{S_n^{*}/\sqrt{n}}=\frac{\bar{X}-\mu}{S_n/\sqrt{n-1}} \sim t(n-1).$$

证 $\because$ $U=\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$，$V=\dfrac{(n-1)S_n^{*2}}{\sigma^2} \sim \chi^2(n-1)$，

且两者独立，由 t 分布的定义知

$$T=\frac{U}{\sqrt{\dfrac{V}{n-1}}}=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\Bigg/\sqrt{\frac{(n-1)S_n^{*2}}{\sigma^2(n-1)}}=\frac{\bar{X}-\mu}{S_n^{*}/\sqrt{n}}$$

$$=\frac{\bar{X}-\mu}{S_n/\sqrt{n-1}} \sim t(n-1).$$

例1 设X和$Y_1,\cdots,Y_n$分别来自正态总体$N(0,\sigma_1^2)$ $N(\mu_2,\sigma_2^2)$且相互独立的样本，试求

$$F=\frac{X^2\sigma_2^2}{S_n^{*2}\sigma_1^2},\quad 其中S_n^{*2}=\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2$$

的概率分布, 并写出分布参数.

解 由卡方分布的定义有

$$\frac{X^2}{\sigma_1^2}\sim\chi^2(1)$$

$$\frac{(n-1)S_n^{*2}}{\sigma_2^2}=\sum_{i=1}^{n}\left(\frac{Y_i-\bar{Y}}{\sigma_2}\right)^2\sim\chi^2(n-1)$$

又因为X^2与$\dfrac{(n-1)S_n^{*2}}{\sigma_2^2}$相互独立，

由F分布性质知

$$\boxed{F=\frac{X^2\sigma_2^2}{S_n^{*2}\sigma_1^2}}=\frac{\dfrac{X^2}{\sigma_1^2}/1}{\dfrac{(n-1)S_n^{*2}}{\sigma_2^2}/(n-1)}$$

$$\boxed{\sim F(1,n-1).}$$

例2 某厂生产的灯泡使用寿命$X \sim N(2250, 250^2)$现进行质量检查，方法如下：任意挑选若干个灯泡，如果这些灯泡的平均寿命超过2200h，就认为该厂生产的灯泡质量合格，**若要使通过检验的概率超过0.997，问至少检查多少只灯泡?**

解 以$\overline{X}$记样本均值，则 $\overline{X} \sim N(2250, \dfrac{250^2}{n})$

$$P(\text{合格}) = P(\overline{X} > 2200)$$

$$= P\left(\frac{\overline{X} - 2250}{250/\sqrt{n}} > \frac{2200 - 2250}{250/\sqrt{n}}\right)$$

$$= 1 - \Phi\left(\frac{2200 - 2250}{250/\sqrt{n}}\right) > 0.997$$

即

$$1-\Phi(-\frac{\sqrt{n}}{5})$$

$$=\Phi(\frac{\sqrt{n}}{5})>0.997$$

$$\Rightarrow\frac{\sqrt{n}}{5}>u_{0.003}=2.75\Rightarrow n\geq 190$$

所以，要是检查能通过的概率超过0.997，至少应该检查190只灯泡.

2. 样本来自两个正态总体

定理5.4 若$X \sim N(\mu_1,\sigma_1^2)$, $Y \sim N(\mu_2,\sigma_2^2)$,

X与Y相互独立. 样本$(X_1, X_2, \cdots, X_{n_1})$ 与$(Y_1, Y_2, \cdots, Y_{n_2})$分别来自总体$X$和$Y$，则

$$(1)\quad \bar{X}-\bar{Y} \sim N\left(\mu_1-\mu_2, \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)$$

或

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

例3 设$(X_1, X_2, \cdots, X_{10})$和$(Y_1, Y_2, \cdots, Y_{15})$是来自总体$N(20,3)$的两个独立的样本，求

$$P\{|\overline{X}-\overline{Y}|>0.3\}.$$

解 $$\overline{X}=\frac{1}{10}\sum_{i=1}^{10}X_i \sim N(20, \frac{3}{10}),$$

$$\overline{Y}=\frac{1}{15}\sum_{i=1}^{15}Y_i \sim N(20, \frac{3}{15}),$$

$$\therefore \quad \overline{X}-\overline{Y} \sim N(0, \frac{3}{10}+\frac{3}{15})=N(0, \frac{1}{2}),$$

故 $\dfrac{\overline{X}-\overline{Y}}{\sqrt{\dfrac{1}{2}}} \sim N(0,1)$

从而 $P\{|\overline{X}-\overline{Y}|>0.3\}=1-P\{|\overline{X}-\overline{Y}|\le 0.3\}$

$$=1-P\left\{\left|\frac{\overline{X}-\overline{Y}}{\sqrt{\frac{1}{2}}}\right|\le\frac{0.3}{\sqrt{\frac{1}{2}}}\right\}=1-[\Phi(0.3\sqrt{2})-\Phi(-0.3\sqrt{2})]$$

$\boxed{\because \Phi(-0.3\sqrt{2})=1-\Phi(0.3\sqrt{2})}$

$$=2[1-\Phi(0.3\sqrt{2})]\approx 2(1-0.6628)=0.6744.$$

(2) 当 $\sigma_1^2=\sigma_2^2=\sigma^2$ 未知时，

由定理5.4(1)，知 $\bar{X}-\bar{Y}\sim N(\mu_1-\mu_2,\dfrac{\sigma^2}{n_1}+\dfrac{\sigma^2}{n_2})$

$$\therefore\quad U=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim N(0,1),$$

又由 $\dfrac{(n_1-1)S_1^{*2}}{\sigma^2}\sim\chi^2(n_1-1),\ \dfrac{(n_2-1)S_2^{*2}}{\sigma^2}\sim\chi^2(n_2-1),$

且它们相互独立, 故由 χ^2 分布的可加性知

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

由于 U 与 V 相互独立,按 t 分布的定义

$$T=\frac{U}{\sqrt{V/(n_1+n_2-2)}}$$

$$=\frac{\left[(\bar{X}-\bar{Y})-(\mu_1-\mu_2)\right]/\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}{\sqrt{((n_1-1)S_1^{*2}+(n_2-1)S_2^{*2})/\sigma^2(n_1+n_2-2)}}$$

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

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

$\therefore$ 当 $\sigma_1^2 = \sigma_2^2 = \sigma^2$ 未知时，

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

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

$$= \sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}}.$$

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

$\dfrac{(n_1-1)S_1^{*2}}{\sigma_1^2}\sim\chi^2(n_1-1),$ $\dfrac{(n_2-1)S_2^{*2}}{\sigma_2^2}\sim\chi^2(n_2-1),$

由假设 S_1^{*2}, S_2^{*2} 独立，则由 F 分布的定义知

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

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

例3 设$X_1,X_2,\cdots,X_n$是来自正态总体$N(\mu,\sigma^2)$的样本

记 $$\boxed{Y_1=\frac{1}{6}\sum_{i=1}^{6}X_i,Y_2=\frac{1}{3}\sum_{i=7}^{9}X_i}$$

$$S^2=\frac{1}{2}\sum_{i=7}^{9}(X_i-Y_2)^2 \qquad Z=\frac{\sqrt{2}(Y_1-Y_2)}{S}$$

试证明：$Z\sim t(2)$.

解 $\because Y_1=\overline{X}_A\sim N(\mu,\frac{\sigma^2}{6}),Y_2=\overline{X}_B\sim N(\mu,\frac{\sigma^2}{3})$

且Y_1,Y_2相互独立

所以 $Y_1 - Y_2 \sim N(0, \dfrac{\sigma^2}{2})$

$$Z = \frac{\sqrt{2}(Y_1 - Y_2)}{S}$$

从而有 $\dfrac{\sqrt{2}(Y_1 - Y_2)}{\sigma} \sim N(0,1)$

$$S^2 = \frac{1}{2}\sum_{i=7}^{9}(X_i - Y_2)^2 = \frac{1}{2}\sum_{i=7}^{9}(X_i - \overline{X}_B)^2$$

又因为 $\dfrac{2S^2}{\sigma^2} \sim \chi^2(2)$, 且 $\dfrac{\sqrt{2}(Y_1 - Y_2)}{\sigma}$ 与 $\dfrac{2S^2}{\sigma^2}$ 独立

所以 $$Z = \frac{\sqrt{2}(Y_1 - Y_2)/\sigma}{\sqrt{2S^2/2\cdot\sigma^2}} = \frac{\sqrt{2}(Y_1 - Y_2)}{S} \sim t(2).$$

例4 设 $X_1, X_2, \cdots, X_n$ 是来自正态总体 $N(\mu, \sigma^2)$ 的样本，$\bar{X}$ 和 S_n^2 分别为样本均值与方差,又设 $X_{n+1} \sim N(\mu, \sigma^2)$ 且与 $X_1, X_2, \cdots, X_n$ 相互独立，试求常数 C 使得 $F = C(X_{n+1} - \bar{X})^2 / S_n^2$ 服从 $F(1, n-1)$.

解 因为 $\bar{X} \sim N(\mu, \frac{\sigma^2}{n}), X_{n+1} \sim N(\mu, \sigma^2)$

所以，由正态分布的线性性得

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

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

从而有 $$U=\left[\frac{(X_{n+1}-\bar{X})}{\sigma}\cdot\sqrt{\frac{n}{n+1}}\right]^2\sim\chi^2(1)$$

另一方面，有样本方差的性质知

$$V=\frac{nS_n^2}{\sigma^2}\sim\chi^2(n-1)$$ 且 $\bar{X}$ 与 S_n^2 相互独立

$\therefore U$ 与 V 相互独立

所以 $C=(n-1)/(n+1).$

1)

内容小结

抽样分布定理

1、单正态总体的抽样分布 $X \sim N(\mu, \sigma^2)$,

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

$$(2)\ U = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$(3)\ \frac{nS_n^2}{\sigma^2} = \frac{(n-1)S_n^{*2}}{\sigma^2} \sim \chi^2(n-1)$$

$$(4)\ \frac{\bar{X}-\mu}{S_n^*/\sqrt{n}} = \frac{\bar{X}-\mu}{S_n/\sqrt{n-1}} \sim t(n-1).$$

2 两正态总体的抽样分布

若$X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$,

(1) $\bar{X} \sim N(\mu_1, \frac{\sigma_1^2}{n_1})$　　$\bar{Y} \sim N(\mu_2, \frac{\sigma_2^2}{n_2})$

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

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

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sigma(\frac{1}{n_1} + \frac{1}{n_2})} \sim N(0,1)$$

$$\frac{(n_1-1)S_1^{*2}}{\sigma_1^2}+\frac{(n_2-1)S_2^{*2}}{\sigma_2^2}\sim\chi^2(n_1+n_2-1)$$

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

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

$$(3)\ \frac{(n_1-1)S_1^{*2}}{\sigma_1^2} \sim \chi^2(n_1-1), \quad \frac{(n_2-1)S_2^{*2}}{\sigma_2^2} \sim \chi^2(n_2-1),$$

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

5-3 抽样分布

Thank You !