06

数理统计的基本概念

引例　截尾试验

从一大批产品中逐件随机抽取检查，

(1) 如果发现废品就停止检查，认为该批产品不合格；

(2) 若抽查到第n_0件仍未发现废品也停止检查，认为该批产品合格.

设产品的废品率为p，问平均要检查多少件产品？

解 设 X 为所检查产品的件数，则

$$P(X=k)=(1-p)^{k-1}p\,,\quad k=1,2,\dots,n_0-1,\qquad P(X=n_0)=(1-p)^{n_0-1}\,.$$

$$E(X)=\sum_{k=1}^{n_0-1}k(1-p)^{k-1}p+n_0(1-p)^{n_0-1}$$

$$=p\Big(\sum_{k=1}^{n_0-1}q^k\Big)'+n_0q^{n_0-1}\qquad(q=1-p)$$

$$=p\Big(\frac{1-q^{n_0}}{1-q}-1\Big)'+n_0q^{n_0-1}=\frac{1-(1-p)^{n_0}}{p}$$

引例　截尾试验

概率问题

p已知，X为检查件数，则

$$P(X=k)=\begin{cases}(1-p)^{k-1}p\,, & k=1,2,\cdots,n_0-1,\\ (1-p)^{k-1}, & k=n_0\,.\end{cases}$$

$$E(X)=\sum_{k=1}^{n_0-1} k\,(1-p)^{k-1}p+n_0\,(1-p)^{n_0-1}=\frac{1-(1-p)^{n_0}}{p}$$

统计问题

p未知，确定适当的n_0,

若$X<n_0$，则认为 $p>p_0$（不合格）；

若$X=n_0$，则认为 $p\le p_0$（合格）.

总体与样本

总体

研究对象的全体 $\xrightarrow{\text{量化}}$ 指标集 $\xrightarrow{\text{规律}}$ R.V. X 或$F(x)$

样本

总体的部分个体：$X_1, X_2, \cdots, X_n$ 独立同分布于$F(x)$

试验前：$X_1, X_2, \cdots, X_n$为R.V.

试验后：$x_1, x_2, \cdots, x_n$为样本观察值

n：样本容量（样本大小）

总体与样本

数理统计的基本思想

由**样本**对**总体**的分布（特征）进行合理地**推断**

06

数理统计的基本概念

理论分布函数

理论分布函数$F(x)$

对总体$F(x)$：样本的联合分布函数 $F(x_1, x_2, \cdots, x_n)=\prod_{i=1}^{n} F(x_i)$

对离散型总体：$P(X_1=x_1, X_2=x_2, \cdots, X_n=x_n)=\prod_{i=1}^{n} P(X=x_i)$

如： $X_1, X_2, \cdots, X_n$ 为取自总体 $B(1, p)$ 的样本，则其联合分布律

$$P(X_1=x_1, X_2=x_2, \cdots, X_n=x_n)=\prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

理论分布函数

对离散型总体：$P(X_1=x_1, X_2=x_2, \cdots, X_n=x_n)=\prod_{i=1}^{n} P(X_i=x_i)$

如：$X_1, X_2, \cdots, X_n$ 为取自总体 $B(1,p)$ 的样本，则其联合分布律

$$P(X_1=x_1, X_2=x_2, \cdots, X_n=x_n)=\prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}=p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

对连续型总体：$f(x_1, x_2, \cdots, x_n)=\prod_{i=1}^{n} f(x_i)$

如：$X_1, X_2, \cdots, X_n$ 为取自总体 $N(\mu,1)$ 的样本，则其联合密度函数

$$f(x_1, x_2, \cdots, x_n)=\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2}}=(2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

06

数理统计的基本概念

经验分布函数

经验分布函数 $F_n(x)$

(1) $x_1, x_2, \cdots, x_n \Longrightarrow x_1^* \le x_2^* \le \cdots \le x_n^*$

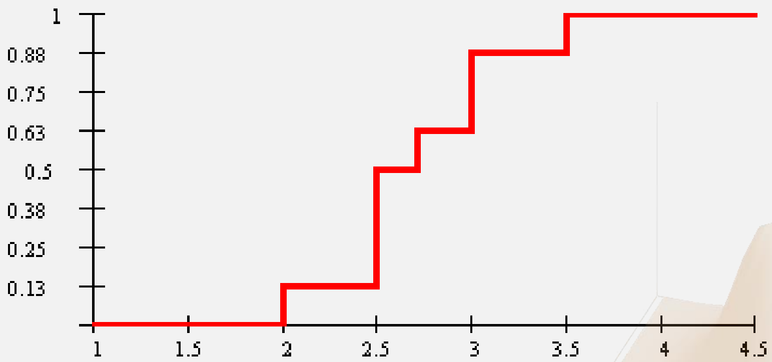
(2) $$F_n(x) = \begin{cases} 0\,, & x < x_1^*, \\ k/n, & x_k^* \le x < x_{k+1}^*, \quad k=1,2,\cdots,n-1, \\ 1, & x \ge x_n^*. \end{cases}$$

经验分布函数

例 随机地观测总体X 得8个数据：2.5, 3, 2.5, 3.5, 3, 2.7, 2.5, 2, 试求X 的一个经验分布函数.

解 $2 < 2.5 = 2.5 = 2.5 < 2.7 < 3 = 3 < 3.5$

$$F_8(x) = \begin{cases} 0, & x < 2, \\ 1/8, & 2 \le x < 2.5, \\ 4/8, & 2.5 \le x < 2.7, \\ 5/8, & 2.7 \le x < 3, \\ 7/8, & 3 \le x < 3.5, \\ 1, & x \ge 3.5. \end{cases}$$

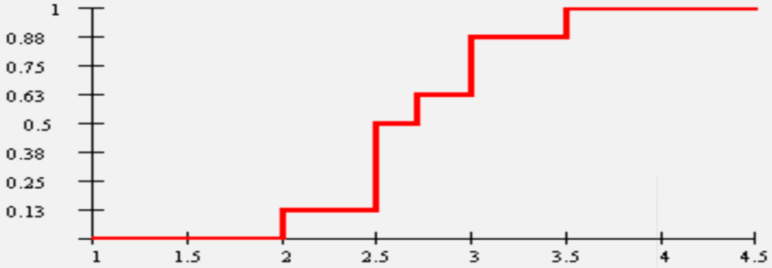


经验分布函数

例 随机地观测总体X 得8个数据：2.5, 3, 2.5, 3.5, 3, 2.7, 2.5, 2, 试求X 的一个经验分布函数.

解 $2 < 2.5 = 2.5 = 2.5 < 2.7 < 3 = 3 < 3.5$

$$F_8(x) = \begin{cases} 0\,, & x < 2, \\ 1/8, & 2 \le x < 2.5\,, \\ 4/8, & 2.5 \le x < 2.7, \\ 5/8, & 2.7 \le x < 3, \\ 7/8, & 3 \le x < 3.5, \\ 1, & x \ge 3.5. \end{cases}$$

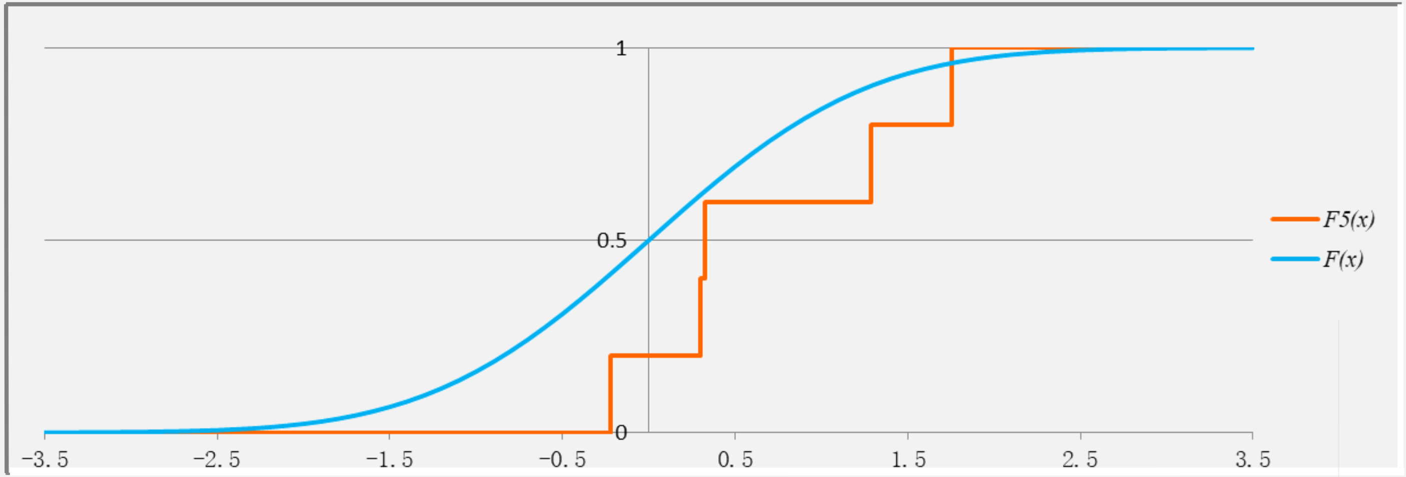


X	2	2.5	2.7	3	3.5
P	1/8	3/8	1/8	2/8	1/8

经验分布函数

格列汶科定理

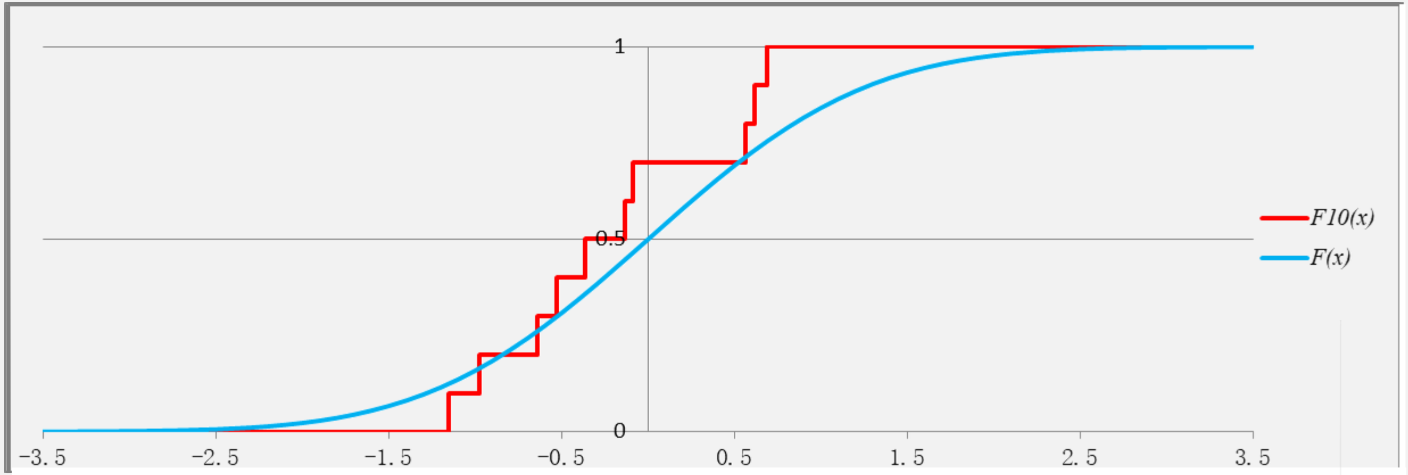
$$P(\lim_{n\to\infty}\sup_{x}|F_n(x)-F(x)|=0)=1$$



经验分布函数

格列汶科定理

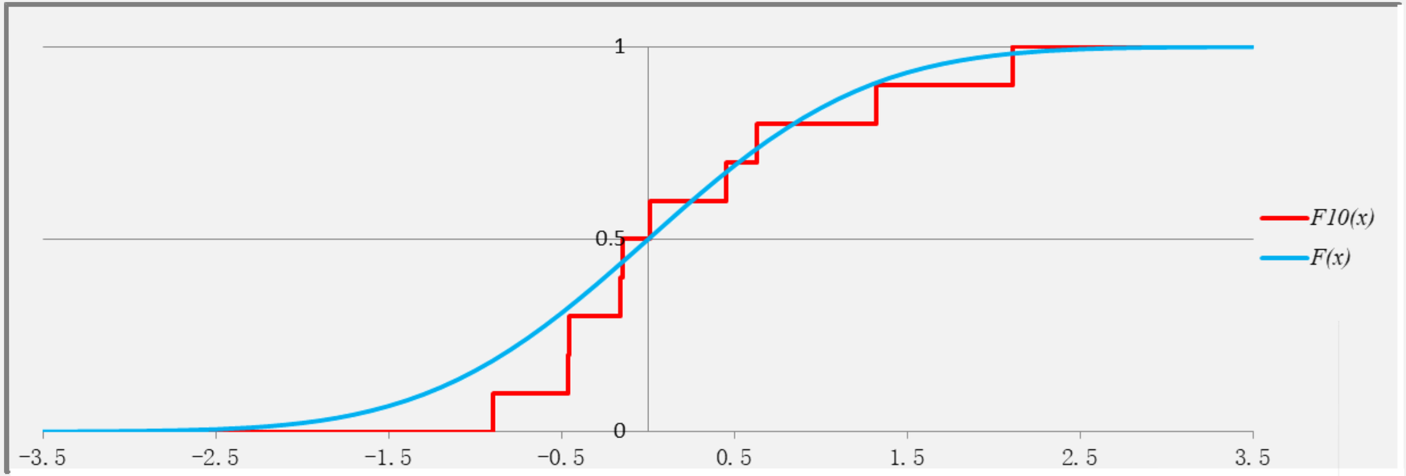
$$P\left(\lim_{n\to\infty}\sup_{x}\left|F_n(x)-F(x)\right|=0\right)=1$$



经验分布函数

格列汶科定理

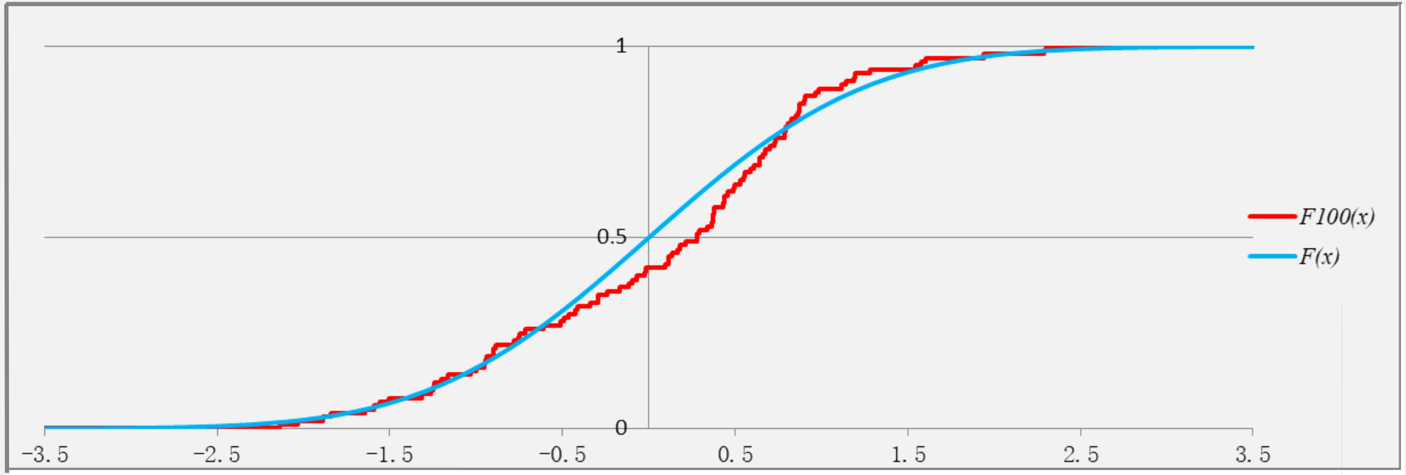
$$P\left(\lim_{n\to\infty}\sup_{x}\left|F_n(x)-F(x)\right|=0\right)=1$$



经验分布函数

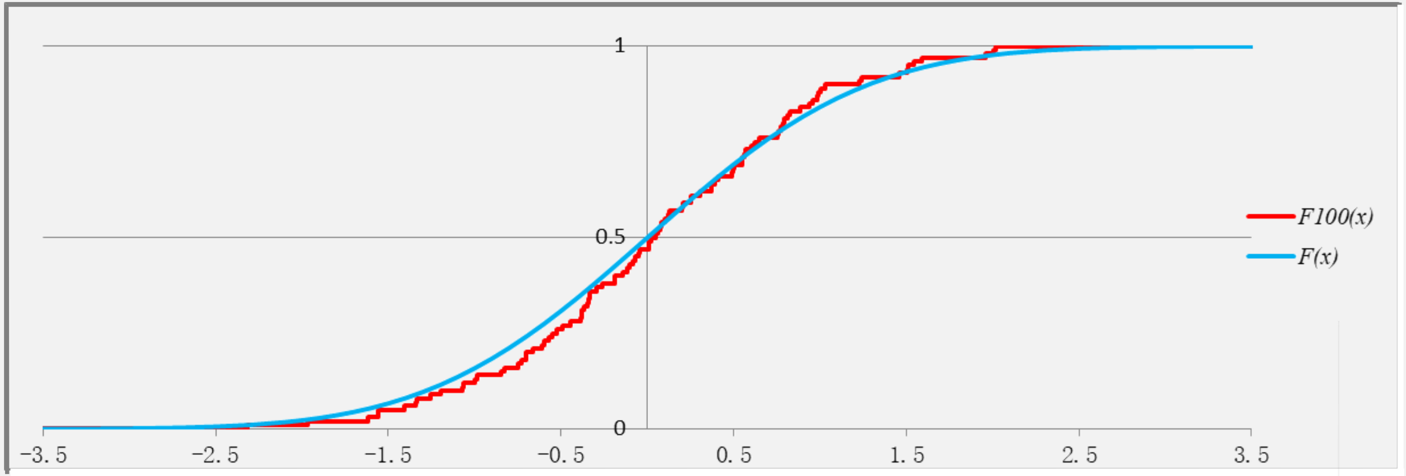
格列汶科定理

$$P\left(\lim_{n\to\infty}\sup_{x}\left|F_n(x)-F(x)\right|=0\right)=1$$



经验分布函数

格列汶科定理 $$P\left(\lim_{n\to\infty}\sup_{x}\left|F_n(x)-F(x)\right|=0\right)=1$$



经验分布函数

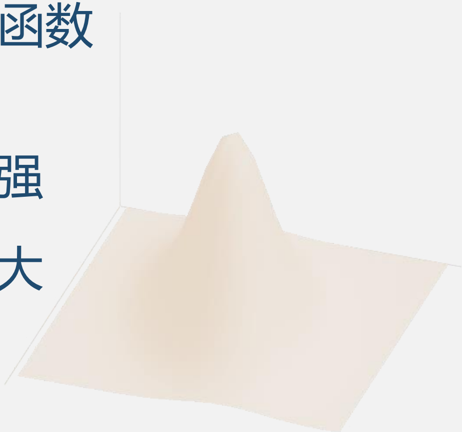
理论分布函数$F(x)$ ： 客观存在的未知分布函数

统计推断的目标

经验分布函数$F_n(x)$ ： 由样本构建的离散型分布函数

优点：计算简单，可视化强

缺点：要求样本容量充分大



06

数理统计的基本概念

统计量

定义 设$X_1, X_2, \cdots, X_n$为来自总体X的一个样本，若

（1）$t = g(x_1, x_2, \cdots, x_n)$ 连续;（$T = g(X_1, X_2, \cdots, X_n)$ 为随机变量即可)

（2）$t = g(x_1, x_2, \cdots, x_n)$ 中不含有关总体的未知参数.

则称 $T = g(X_1, X_2,\cdots, X_n)$ 为统计量，称 $t = g(x_1, x_2, \cdots, x_n)$ 为统计量观察值.

例 设$X_1, X_2, \cdots, X_n$为来自总体$N(\mu, \sigma^2)$的样本，其中μ已知，σ^2未知，则

(1) $X_1 + X_n$ (2) $\frac{1}{n}\sum\limits_{i=1}^{n}(X_i-\mu)^2$ (3) $\sum\limits_{i=1}^{n}\frac{|X_i|}{\sigma}$ (4) $\min\limits_{1\le i\le n}\{X_i\}$

是统计量.

统计量

常用统计量

1、样本均值

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

提取E(X)的信息

2、样本方差

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

提取D(X)的信息

样本标准差

$$S = \sqrt{S^2}$$

3、样本k阶原点矩

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

提取E(X^k)的信息

$$A_1 = \overline{X}$$

4、样本k阶中心矩

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

提取E(X-EX)^{k}的信息

$$B_2 = \frac{n-1}{n} S^2 = \tilde{S}^2$$

06

数理统计的基本概念

统计量

常用统计量

顺序统计量

$X_1, X_2, \cdots, X_n \Rightarrow X_1^* \le X_2^* \le \cdots \le X_n^*$

	X_1	X_2	X_3	X_4	X_5
(1)	3	1	10	5	6
(2)	2	6	7	2	8
(3)	8	3	9	10	5

	X_1^*	X_2^*	X_3^*	X_4^*	X_5^*
(1)	1	3	5	6	10
(2)	2	2	6	7	8
(3)	3	5	8	9	10

统计量

常用统计量

顺序统计量

$X_1, X_2, \cdots, X_n \Rightarrow X_1^* \le X_2^* \le \cdots \le X_n^*$

注意：X_i^* 并不是 $X_1, X_2, \cdots, X_n$ 中的一个，如

$X_1, X_2 \overset{\text{iid}}{\sim} \mathrm{B}(1, p)$

$\Rightarrow$ $X_1^* \sim \mathrm{B}(1, \boldsymbol{p}^2),$

$X_2^* \sim \mathrm{B}(1, \mathbf{1-(1-}\boldsymbol{p}\mathbf{)}^2).$

统计量

常用统计量

顺序统计量

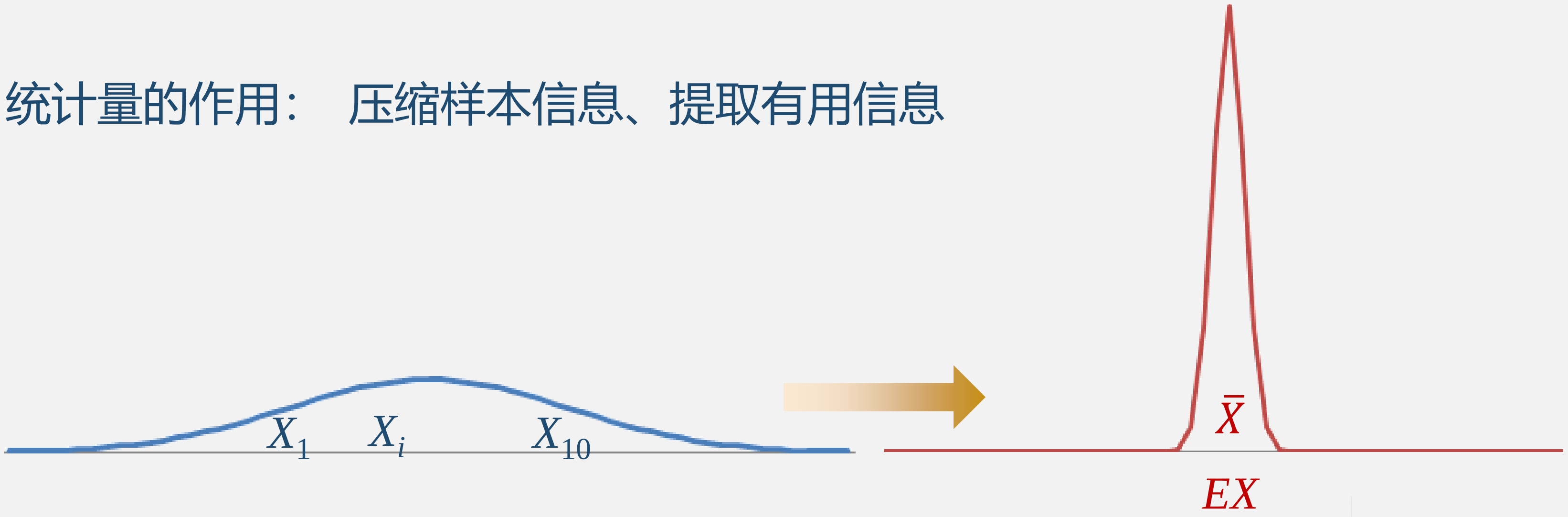
$$X_1, X_2, \cdots, X_n \Rightarrow X_1^* \le X_2^* \le \cdots \le X_n^*$$

样本中位数 $$\tilde{x} = \begin{cases} X_{m+1}^* & n=2m+1 \\ \frac{1}{2}(X_m^* + X_{m+1}^*) & n=2m \end{cases}$$ 提取分布中位数的信息

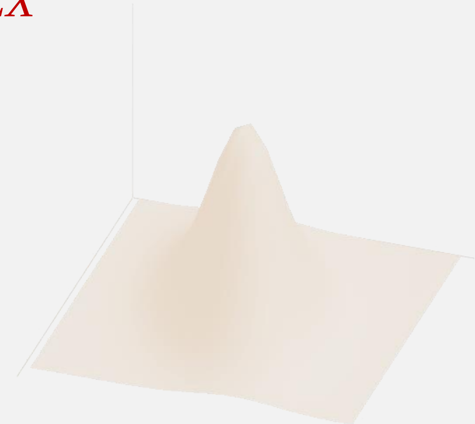
样本 极 差 $$R = X_n^* - X_1^*$$ 提取分布范围的信息

统计量

统计量的作用：　压缩样本信息、提取有用信息



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



06

数理统计的基本概念

χ^2分布

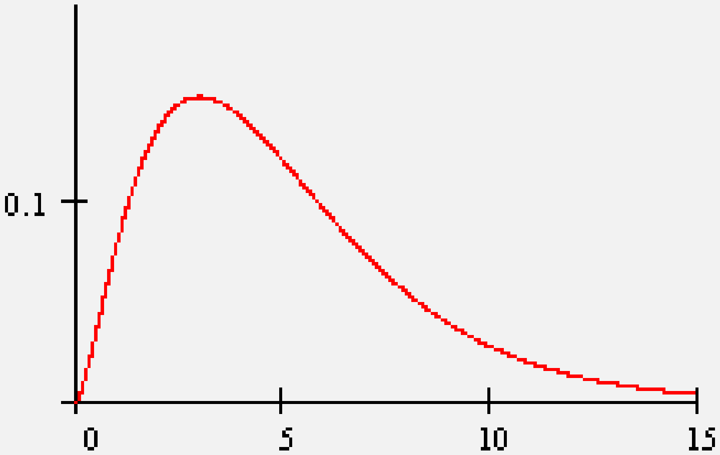
设$X_1, X_2, \cdots, X_n$独立同分布于$N(0,1)$，则称

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

服从自由度为n的χ^2分布，记为$\chi^2 \sim \chi^2(n)$.

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

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$$



χ^2分布

数字特征

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

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

可加性

$\chi_1^2 \sim \chi^2(n_1)$与$\chi_2^2$

上侧分位数

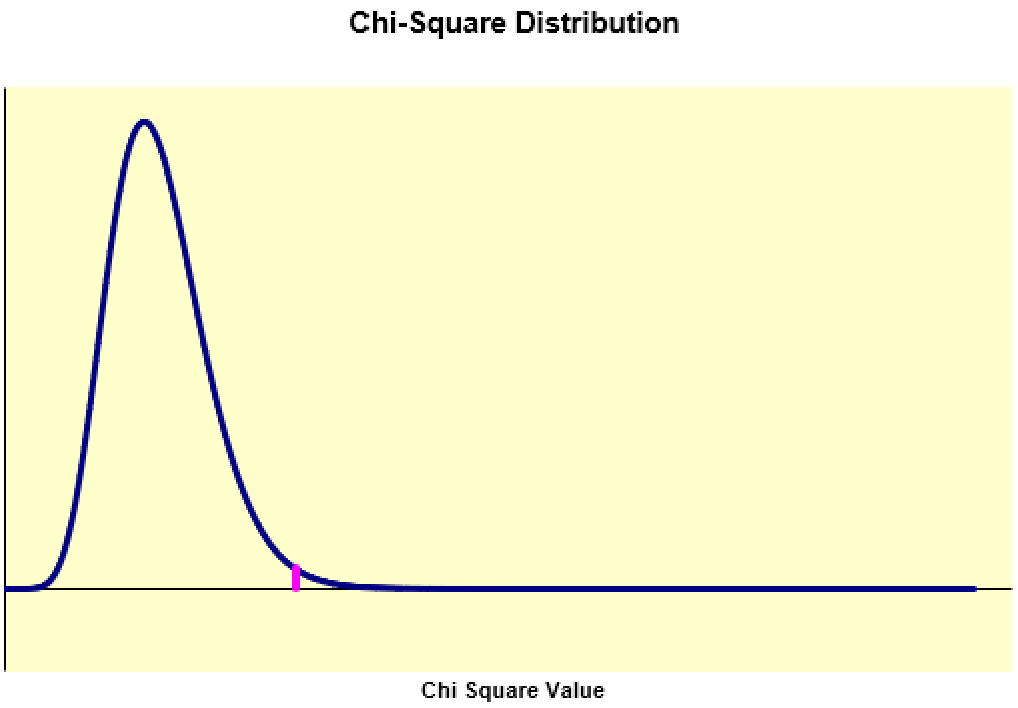
$$P(\chi^2 > \chi_\alpha^2(n)) = \alpha$$

n \ α	0.25	0.10	0.05	0.025	0.01	0.005
14	17.117	21.064	23.685	26.119	29.141	31.319
15	18.245	22.307	24.996	27.488	30.578	32.801
16	19.369	23.542	26.296	28.845	32.000	34.267
17	20.489	24.769	27.587	30.191	33.409	35.718
18	21.605	25.989	28.869	31.526	34.805	37.156
19	22.718	27.204	30.144	32.852	36.191	38.582
20	23.828	28.412	31.410	34.170	37.566	39.997
21	24.953	29.615	32.671	35.479	38.932	41.401

$$\chi_{0.01}^2(20) = 37.566$$

查表

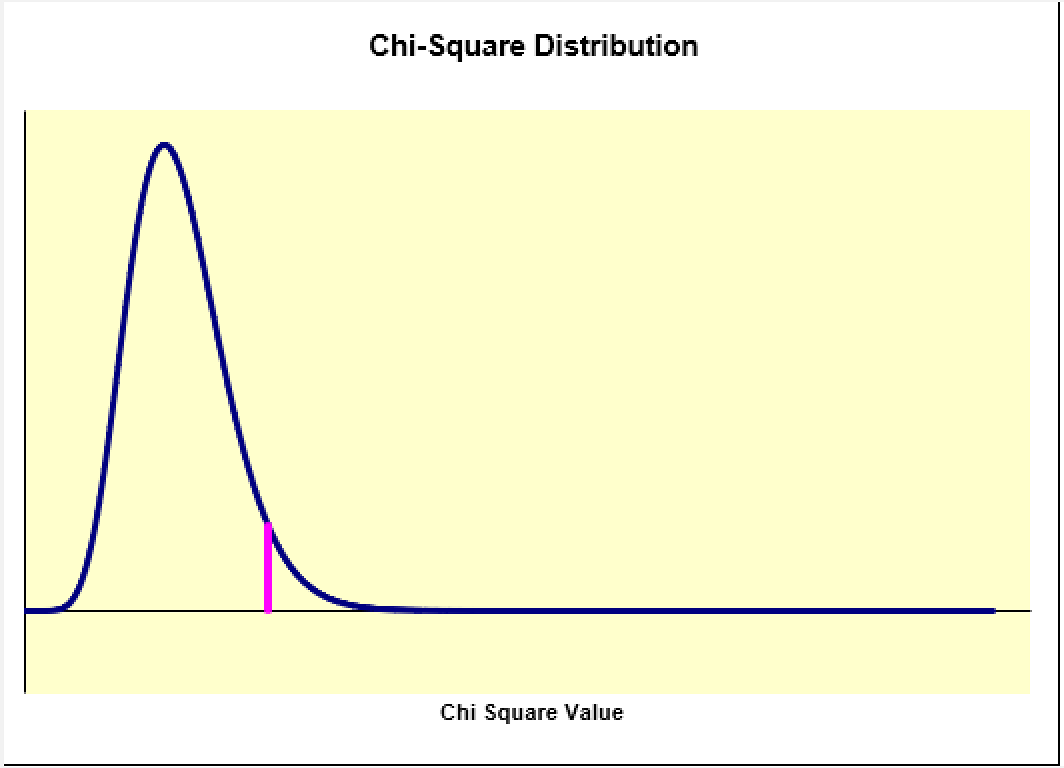
χ^2分布



Degrees of Freedom	20
Alpha Value	1%

Chi-Square Value	37.566

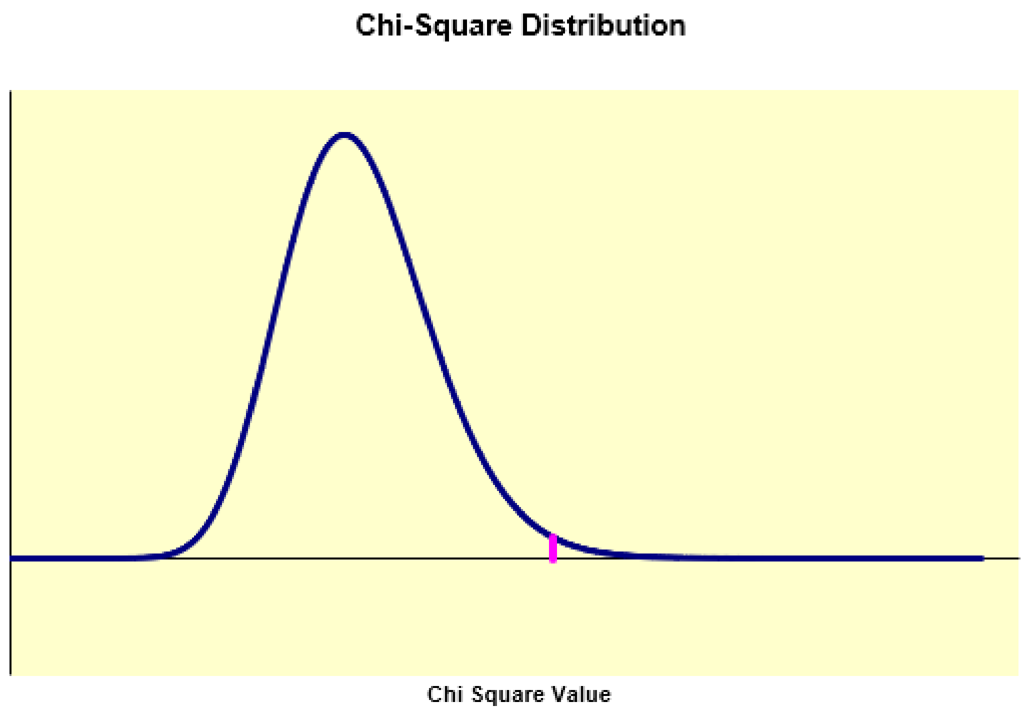
χ^2分布



Degrees of Freedom	20
Alpha Value	5%

Chi-Square Value	31.410

χ^2分布

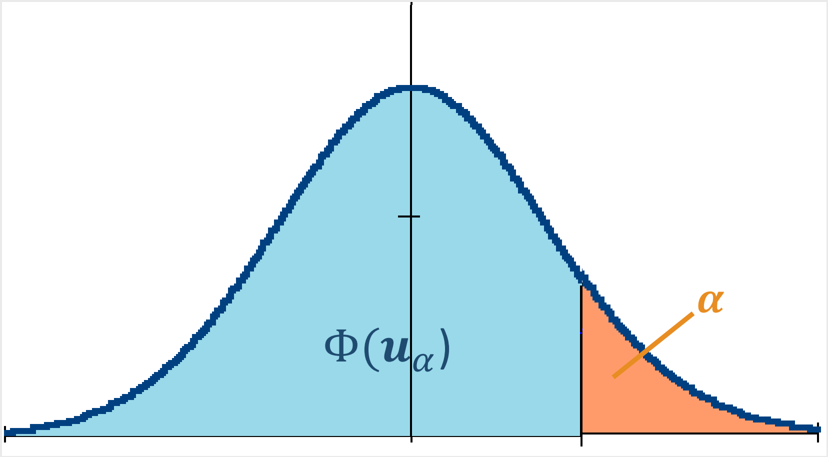


Degrees of Freedom	45
Alpha Value	1%

Chi-Square Value	69.957

χ^2分布

上侧分位数　$\chi_\alpha^2(n) \approx \frac{1}{2}\left(\boldsymbol{u}_\alpha + \sqrt{2n-1}\right)^2 \quad (n > 45)$



$\Phi\ (\boldsymbol{u}_\alpha) = 1 - \alpha$

χ^2分布

例 设 $X_i\,(i=1,2,\dots,n)$ 为总体 $X \sim N\,(0,\,0.04)$的一个样本，求 $P(\sum_{i=1}^{15} X_i^2 > 1.223)$.

解

$$X_i \sim N\,(0,\,0.04),\ i=1,\,2,\,\dots,\,n$$

$$Y_i = X_i\,/0.2 \sim N(0,\,1),\ i=1,\,2,\,\dots,\,n$$

$$\sum_{i=1}^{15} Y_i^2 = \sum_{i=1}^{15} \left(\frac{X_i}{0.2}\right)^2 \sim \chi^2(15)$$

$$P\left(\sum_{i=1}^{15} X_i^2 > 1.223\right) = P\left(\sum_{i=1}^{15} \frac{X_i^2}{0.04} > 30.57\right) \approx 0.01$$

06

数理统计的基本概念

t 分布

设 $X\sim N(0,1)$ 与 $Y\sim\chi^2(n)$ 独立，则称 $T=\dfrac{X}{\sqrt{Y/n}}$

服从自由度为 n 的 t 分布，记 $T\sim t(n)$.

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

t 分布

数字特征

$$T \sim t(n)$$

$$n > 1, \qquad E(T) = 0.$$

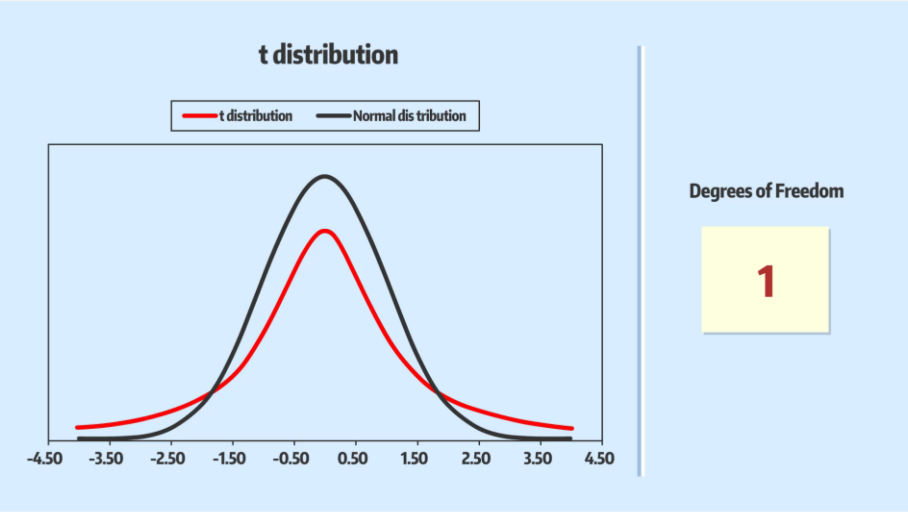
$$n = 1, \qquad f(x) = \frac{1}{\pi(1+x^2)}$$

柯西分布

t 分布

重尾分布

渐近正态性



t 分布

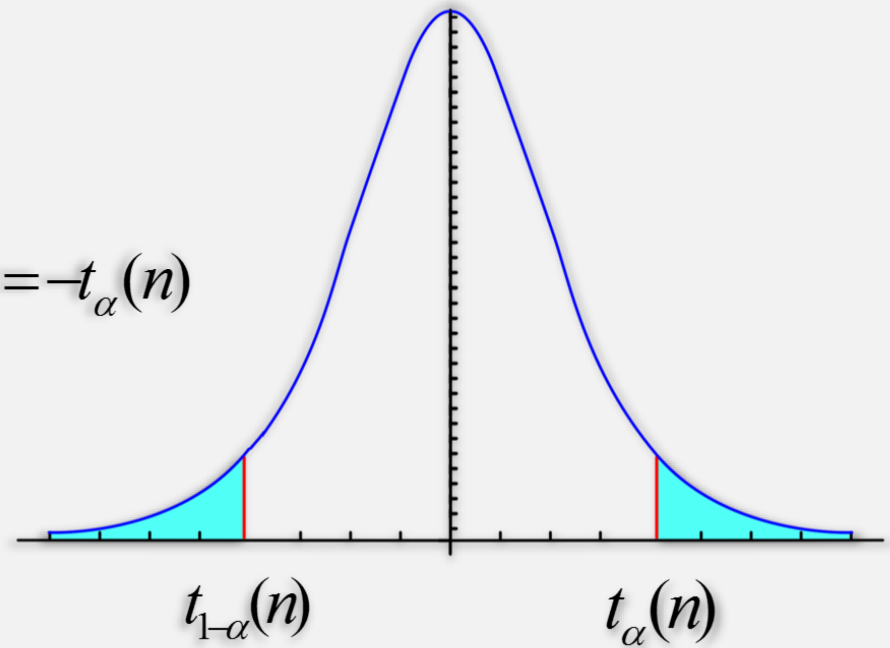
上侧分位数 $P(T > t_\alpha(n)) = \alpha$ $t_{0.05}(6) = 1.9432$

n \ α	0.25	0.10	0.05	0.025	0.01	0.005
1	1.0000	3.0777	6.3138	12.7062	31.8207	63.6574
2	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8409
4	0.7407	1.5332	2.1318	2.7764	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0322
6	0.7176	1.4398	**1.9432**	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9980	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554

查表

t 分布

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



06

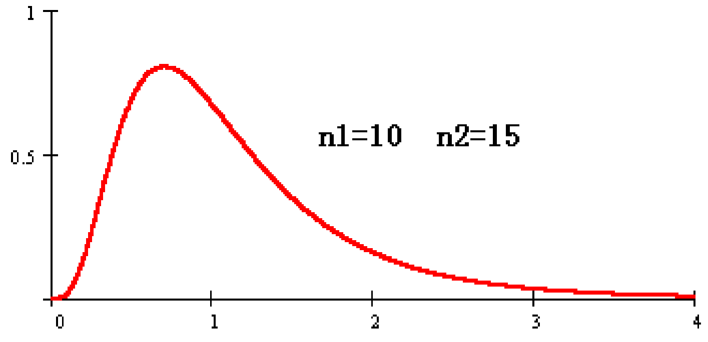
数理统计的基本概念

F 分布

设$X \sim \chi^2(n_1)$与$Y \sim \chi^2(n_2)$相互独立，则称 $F = \dfrac{X/n_1}{Y/n_2}$

服从自由度为(n_1, n_2)的F分布，记为$F \sim F(n_1, n_2)$.

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



上侧分位数

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

F 分布

查表 $F_{0.975}(15,10) = \dfrac{1}{F_{0.025}(10,15)} = \dfrac{1}{3.06} = 0.33$

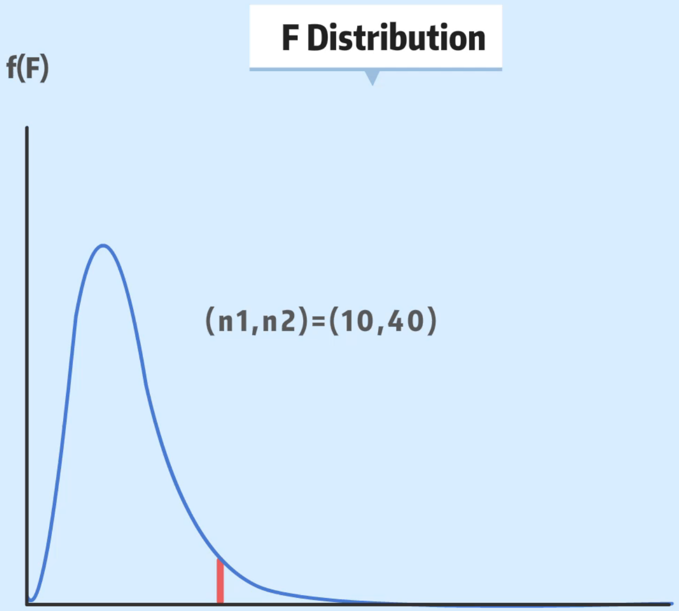
$$\frac{1}{F} = \frac{Y/n_2}{X/n_1} \sim F(n_2, n_1)$$

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

$\alpha = 0.025$

$n_2 \backslash n_1$	8	9	10	12	15	20
13	3.39	3.31	3.25	3.15	3.05	2.95
14	3.29	3.21	3.15	3.05	2.95	2.84
15	3.20	3.12	3.06	2.96	2.86	2.76
16	3.12	3.05	2.99	2.89	2.79	2.68
17	3.06	2.98	2.92	2.82	2.72	2.62
18	3.01	2.93	2.87	2.77	2.67	2.56

F 分布



Numerator Degrees of Freedom	10
Denominator Degrees of Freedom	40
Alpha Value	0.05

F Value	2.0662

F 分布

例 $X \sim N(0,\sigma^2), X_1, X_2, X_3, X_4 \ i.i.d.$

$$Y = (X_1 + X_2)^2 / (X_3 - X_4)^2 \sim ?$$

解 $X_1 + X_2, \ X_3 - X_4 \ \ i.i.d. \ \ N(0, 2\sigma^2)$

$$\chi_1^2 = [(X_1 + X_2)/\sqrt{2}\sigma]^2 \sim \chi^2(1)$$

$$\chi_2^2 = [(X_3 - X_4)/\sqrt{2}\sigma]^2 \sim \chi^2(1)$$

$$\frac{\chi_1^2/1}{\chi_2^2/1} = \frac{(X_1 + X_2)^2}{(X_3 - X_4)^2} = Y \sim F(1,1)$$

06

随机变量及其分布

单正态总体的抽样分布

抽样定理 设$X_1, X_2, \ldots, X_n$为来自正态总体$N(\mu, \sigma^2)$的一个样本，则

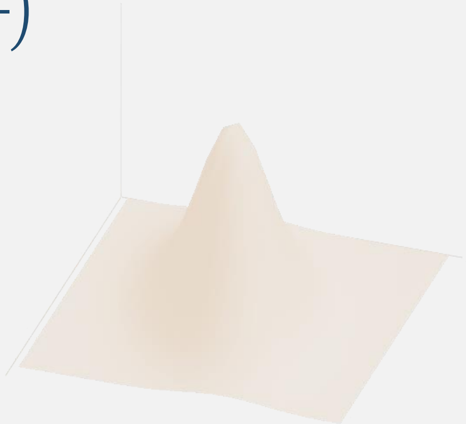
(1) $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$; (2) $\frac{n-1}{\sigma^2}S^2 = \sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma}\right)^2 \sim \chi^2(n-1)$;

(3) $\bar{X}$与S^2相互独立.

证明(1) $$\bar{X} = \sum_{i=1}^{n}\frac{1}{n}X_i \sim N\left(\sum_{i=1}^{n}\frac{1}{n}\mu_i, \sum_{i=1}^{n}\frac{1}{n^2}\sigma_i^2\right) \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

说明(2) $$\sum_{i=1}^{n}(X_i - \bar{X}) = \sum_{i=1}^{n}X_i - \bar{X} = 0$$

解释(3) $\bar{X} \longrightarrow \mu$ $\quad S^2 \longrightarrow \sigma^2$



单正态总体的抽样分布

例 设$X_1, X_2, \cdots, X_n$为来自正态总体$N(\mu, \sigma^2)$的一个样本， 其中σ^2已知，用样本均值$\bar{X}$估计μ，为了有95%的把握保证估计误差小于0.01，样本容量n应该取多少?

解

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

$$P(|\bar{X} - \mu| < 0.01) = P(\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| < 0.01\frac{\sqrt{n}}{\sigma})$$

$$0.95 = 2\Phi\left(\frac{\sqrt{n}}{100\sigma}\right) - 1$$

$$\rightarrow \Phi\left(\frac{\sqrt{n}}{100\sigma}\right) = 0.975 \quad \rightarrow \frac{\sqrt{n}}{100\sigma} = u_{0.025} = 1.96 \quad \rightarrow n = (196\sigma)^2$$

σ^2未知?

06

数理统计的基本概念

正态总体的抽样分布

抽样定理 设$X_1, X_2, \cdots, X_n$为来自正态总体$N(\mu, \sigma^2)$ 的一个样本，则

(1) $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$； (2) $\frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$； (3)$\bar{X}$与$S^2$相互独立.

推论1 $\boxed{T = \frac{\bar{X}-\mu}{S}\sqrt{n} \sim t(n-1)}$

证明

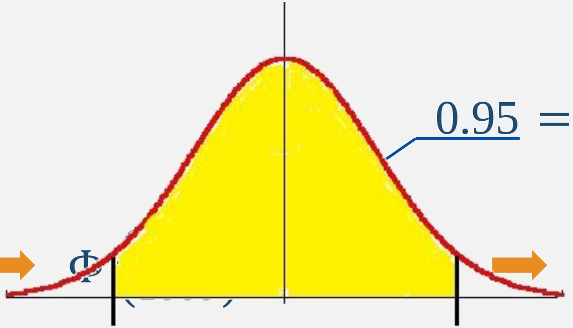
$$\frac{(\bar{X}-\mu)\Big/\sqrt{\frac{\sigma^2}{n}}}{\sqrt{\frac{n-1}{\sigma^2}S^2/(n-1)}} \sim t(n-1)$$

正态总体的抽样分布

例 设$X_1, X_2, \cdots, X_n$为来自正态总体$N(\mu, \sigma^2)$的一个样本， 其中σ^2已知，用样本均值$\bar{X}$估计μ，为了有95%的把握保证估计误差小于0.01，样本容量n应该取多少?

解 由抽样定理 $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$，按题设要求

$$P(|\bar{X}-\mu| < 0.01) = P\left(\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| < 0.01\frac{\sqrt{n}}{\sigma}\right)$$



$$0.95 = 2\Phi\left(\frac{\sqrt{n}}{100\sigma}\right) - 1$$

$$\frac{\sqrt{n}}{100\sigma} = u_{0.025} = 1.96$$

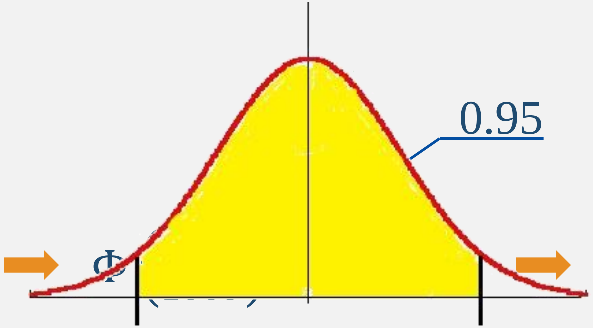
$$n = (196\sigma)^2$$

正态总体的抽样分布

例 设$X_1, X_2, \cdots, X_n$为来自正态总体$N(\mu, \sigma^2)$的一个样本， 其中σ^2未知，用样本均值$\bar{X}$估计μ，为了有95%的把握保证估计误差小于0.01，样本容量n应该取多少?

解 由推论1 $T = \dfrac{\bar{X} - \mu}{S}\sqrt{n} \sim t(n-1)$，按题设要求

$$P(|\bar{X} - \mu| < 0.01) = P\left(|T| < 0.01\frac{\sqrt{n}}{S}\right)$$



$$t_{0.025}(n-1) = \frac{\sqrt{n}}{100S}$$

求解 $n =$?

06

数理统计的基本概念

双正态总体的抽样分布

抽样定理 设$X_1, X_2, \cdots, X_n$为来自正态总体$N(\mu, \sigma^2)$的一个样本，则

(1) $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$； (2) $\frac{n-1}{\sigma^2}S^2 \sim \chi^2(n-1)$； (3) $\bar{X}$与S^2相互独立.

推论1 $T = \frac{\bar{X}-\mu}{S}\sqrt{n} \sim t(n-1)$

推论2 设 $X_1, X_2, \cdots, X_{n_1} \overset{iid}{\sim} N(\mu_1, \sigma_1^2)$： $\bar{X}$， S_1^2

独立样本 $Y_1, Y_2, \cdots, Y_{n_2} \overset{iid}{\sim} N(\mu_2, \sigma_2^2)$： $\bar{Y}$， S_1^2 则

(1) $F = \frac{S_1^2/{\sigma_1}^2}{S_2^2/{\sigma_2}^2} \sim F(n_1-1, n_2-1)$ (2) $\frac{1}{n_1}\sum_{i=1}^{n_1}\left(\frac{X_i-\mu_1}{\sigma_1}\right)^2 \Big/ \frac{1}{n_2}\sum_{i=1}^{n_2}\left(\frac{Y_i-\mu_2}{\sigma_2}\right)^2 \sim F(n_1, n_2)$

双正态总体的抽样分布

推论2 设 $X_1, X_2, \cdots, X_{n_1} \overset{iid}{\sim} N(\mu_1, \sigma_1^2)$: $\bar{X}$， S_1^2

独立样本 $Y_1, Y_2, \cdots, Y_{n_2} \overset{iid}{\sim} N(\mu_2, \sigma_2^2)$: $\bar{Y}$， S_2^2 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) $\dfrac{1}{n_1}\sum_{i=1}^{n_1}\left(\dfrac{X_i-\mu_1}{\sigma_1}\right)^2 \Bigg/ \dfrac{1}{n_2}\sum_{i=1}^{n_2}\left(\dfrac{Y_i-\mu_2}{\sigma_2}\right)^2 \sim F(n_1, n_2)$

证 由抽样定理，及F分布的构造性定义

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

双正态总体的抽样分布

推论3　条件同推论2，且$\sigma_1^2=\sigma_2^2=\sigma^2$，则

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

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

证　由 $\bar X\sim N(\mu_1,\frac{\sigma^2}{n_1})$ 与 $\bar Y\sim N(\mu_2,\frac{\sigma^2}{n_2})$ 相互独立，得 $\bar X-\bar Y\sim N(\mu_1-\mu_2,\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2})$

标准化变量 $$U=\frac{\bar X-\bar Y-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim N(0,1)$$

$\frac{n_1-1}{\sigma^2}S_1^2\sim\chi^2(n_1-1)$，$\frac{n_2-1}{\sigma^2}S_2^2\sim\chi^2(n_2-1)$ $\xrightarrow[\text{可加性}]{\text{相互独立}}$ $V=\frac{1}{\sigma^2}[(n_1-1)S_1^2+(n_2-1)S_2^2]\sim\chi^2(n_1+n_2-2)$

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

双正态总体的抽样分布

例 分别从方差为20和35的两个独立的正态总体中抽取容量为8和10的两个样本,估计第一个样本方差 S_1^2 不小于第二个样本方差 S_2^2 两倍的概率.

解 题设条件为:$(X_1,X_2,\cdots,X_8)\overset{iid}{\sim} N(\mu_1, 20)$, $(Y_1,Y_2,\cdots,Y_{10})\overset{iid}{\sim} \mathrm{N}(\mu_2, 35)$

由推论2知 $F=\dfrac{S_1^2/20}{S_2^2/35}\sim F(7,9)$, 故所求概率为

$$P(S_1^2\geq 2S_2^2) = P\left(\frac{S_1^2/20}{S_2^2/35}\geq 2\times\frac{35}{20}\right)=P(F\geq 3.5)=0.0423$$

查表有:$F_{0.05}(7,9)=3.29$,$F_{0.025}(7,9)=4.20$,所以

$$0.025 < P(S_1^2\geq 2S_2^2) < 0.05$$