数理统计是研究大量随机现象统计规律的一门数学科学，以概率论为基础：

(1) 收集、整理和分析受到随机性影响的数据

(2) 为随机现象选择和检验数学模型

(3) 推断和预测随机现象的性质、特点和统计规律

(4) 为决策提供依据和建议

§1 总体与样本

一、总体、个体及总体分布

总体：与研究对象的某项数量指标相联系的试验的全部可能的观察值。

个体：总体中的每个元素，即试验的每个可能观察值称为个体。

总体的容量：总体中包含的个体的个数。

有限总体：容量为有限的总体。

无限总体：容量为无限的总体。

返回主目录

例如：某工厂生产的灯泡的寿命是一个总体，每一个灯泡的寿命是一个个体；

某学校男生的身高的全体是一个总体，每个男生的身高是一个个体。

检验某工厂生产的零件是正品还是次品，以“1”表示零件是正品，以“0”表示零件是次品，则总体由一些“0”和“1”组成，其中的“0”或“1”为个体。

设X表示联系总体的数量指标的值，则X是一随机变量，X的所有取值即为总体。

总体分布：随机变量X的分布。

返回主目录

例如：某工厂生产的灯泡的寿命是一个总体，从该厂中任取一只灯泡测试它的寿命，用 X 表示该灯泡的寿命，则 X 是随机变量，且 X 的所有取值为总体。一般 X 服从指数分布，θ参数 为该工厂生产的灯泡的平均寿命，故该总体为指数分布总体。

检验某工厂生产的零件是正品还是次品，以“1”表示零件是正品，以“0”表示零件是次品，则总体由一些“0”和“1”组成，其中的“0”或“1”为个体。引进随机变量

$$X=\begin{cases}1 & \text{任取一件产品是正品}\\ 0 & \text{任取一件产品是次品}\end{cases}\quad ,\text{则总体}X\sim b(1,p)$$

一般，总体的分布未知，或总体的分布已知，但某些参数未知。要对总体进行推断，我们对每个个体研究是不可能的，故须抽出部分个体进行研究。

二、样本、样本分布

样本：从总体中抽出的部分个体。

样本容量：样本中所含个体的个数。

每个个体是一个数，所以每次抽出容量为 n 的样本是 n 个数，但由于每个个体的抽取是随机的，故容量为 n 的样本可看成是 n 个随机变量。

容量为 n 的样本用 $X_1,\cdots,X_n$ 表示，其中每个 $X_i(i=1,\cdots,n)$ 都是 随机变量，要求它们满足以下两个特点：

（1）代表性：每个 $X_i(i=1,\cdots,n)$ 与总体 X 同分布；

（2）独立性：$X_1,\cdots,X_n$ 相互独立，即每个观察结果既不影响其它观察结果，也不受其它观察结果的影响。

则称 $X_1,\cdots,X_n$ 为简单**随机样本**，简称为样本。

抽取简单随机样本的方法：

在相同条件下对总体进行 n 次独立，重复的观察。将观察结果按试验的次序依次记为 $X_1, \cdots, X_n$，就得到一个简单随机样本 $X_1, \cdots, X_n$。

当 n 次观察一经完成，就得到一组实数 $x_1, \cdots, x_n$，它们依次是随机变量 $X_1, \cdots, X_n$ 的观察值，称为样本值。

无限总体 $\longrightarrow$ **无放回抽样**

有限总体 $\longrightarrow$ **有放回抽样**

$\dfrac{n}{N}$ **很小**

返回主目录

样本分布

由定义知：若 $(X_1,\cdots,X_n)$ 为来自总体 X 的一个样本，则 $(X_1,\cdots,X_n)$ 是 n 维随机变量，且 $X_1,\cdots,X_n$ 相互独立，故 $(X_1,\cdots,X_n)$ 的联合分布函数为：

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

若设 X 的概率密度为 f(x)，则 $(X_1,\cdots,X_n)$ 的联合概率密度为：

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

若 X 的分布律为 $P\{X=x\}=p(x)$，则 $(X_1,\cdots,X_n)$ 的联合分布律为：

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

返回主目录

例 1 设总体 $X \sim \exp(\theta)$，$X_1,\cdots,X_n$ 为来自总体 X 的一个样本，则 $(X_1,\cdots,X_n)$ 的联合概率密度为：

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

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

$$=\begin{cases}\dfrac{1}{\theta^n}e^{-\frac{1}{\theta}\sum_{i=1}^{n}x_i} & x_1>0,\cdots,x_n>0\\ 0 & \text{其他}\end{cases}$$

例 2 设总体 $X \sim \pi(\lambda)$，$(X_1,\cdots,X_n)$为来自总体 X 的一个样本，则$(X_1,\cdots,X_n)$的联合分布律为：

$$P\{X_1 = x_1,\cdots,X_n = x_n\} = \prod_{i=1}^{n} P\{X_i = x_i\}$$

$$= \frac{\lambda^{\sum_{i=1}^{n} x_i}}{x_1!\cdots x_n!} e^{-n\lambda}$$

$$\boxed{P\{X = x\} = \frac{\lambda^x}{x!} e^{-\lambda}}$$

返回主目录

§2. 直方图和箱线图

通过试验得到许多的观察值，将这些数据加以整理，借助表格或图形加以描述，以便粗略了解总体的分布。

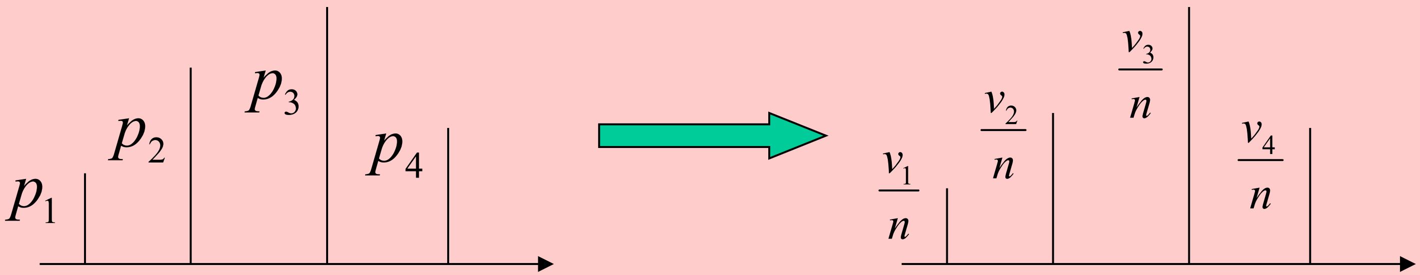
一、直方图

离散型：

设总体X为离散型随机变量，分 布列为$P(X=a_i)=p_i$

令v_i表示事件$\{X=a_i\}$在n次重复独立观测中出现的次数

$$\frac{v_i}{n} \xrightarrow{P} p_i \qquad n \to \infty$$



连续型：

设总体X为连续型随机变量，密 度$f(x)$为未知的

对任一有限区间$[a,b]$，等分成k个子区间，其长度为 $\dfrac{b-a}{k}$

$$a = a_0 < a_1 < \cdots < a_{k-1} < a_k = b$$

v_i表示n次重复独立观测得样本$X_1, X_2, \cdots, X_n$中落在区间$(a_i, a_{i+1}]$中的个数

$$\frac{v_i}{n} \xrightarrow{P} P\{a_i < X \le a_{i+1}\} = \int_{a_i}^{a_{i+1}} f(x)dx$$

当k,n充分大时，　$\frac{v_i}{n} \approx f(a_i)\cdot\frac{b-a}{k}$

定义函数：

当$a_i < x \le a_{i+1}$时，有$f_n(x) = \frac{v_i}{n}\cdot\frac{k}{b-a}$，$i = 0,1,\cdots,m-1$，

称$f_n(x)$在区间$[a,b)$的图形为$[a,b)$上的频率直方图，简称为直方图。

$$f_n(x) \approx f(x)$$

注：直方图中的小矩形的面积等于落在该小区间的频率。

说明：

（1）作直方图时，先取一个区间，其下限比最大的数据稍小，其上限比最大的数据稍大；

（2）将这一区间分为 k 个小区间，通常当 n 较大时 k 取 10~20，当 n<50 时则 k 取 5~6。若 k 取得过大，则会出现某些小区间内频数为零的情况（一般应设法避免）。

（3）分点通常取比数据精度高一位，以避免数据落在分点上。

二、箱线图

定义：设有容量为n的样本观察值$x_1, x_2, \cdots, x_n$，样本p分位数$(0 < p < 1)$记为x_p，它具有以下的性质：(1)至少有np个观察值小于或等于x_p；(2)至少有$n(1-p)$个观察值大于或等于x_p.

将$x_1, x_2, \cdots, x_n$按自小到大的次序排列成$x_{(1)}, x_{(2)}, \cdots, x_{(n)}$

样本p分位数m_p定义为

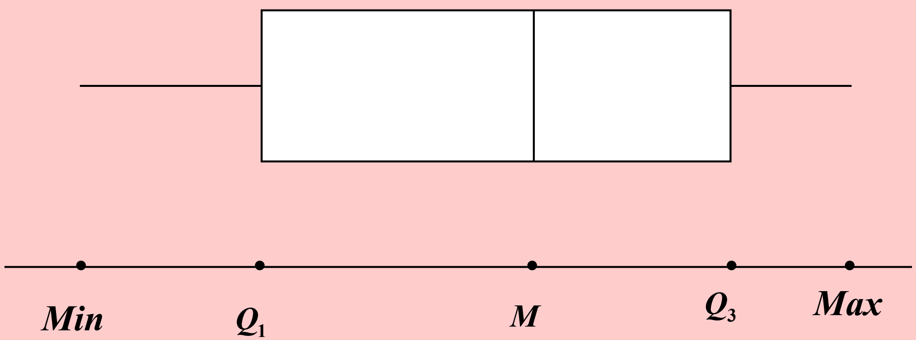
$$x_p = \begin{cases} x_{([np]+1)} & np\text{不是整数} \\ \frac{1}{2}\left[x_{(np)} + x_{(np+1)}\right] & np\text{是整数} \end{cases}$$

特别的，当$p = 0.5$时，样本中位数$x_{0.5}$定义为

$$x_{0.5} = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & n\text{为奇数} \\ \frac{1}{2}\left(X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}\right) & n\text{为偶数} \end{cases}$$

0.25分位数$x_{0.25}$称为第一四分位数，又记为Q_1；0.75分位数$x_{0.75}$称为第三四分位数，记为Q_3。

箱线图的做法：基于5个数据最小值Min，第一四分位数Q_1，中位数M，第三四分位数Q_3和最大值Max.



箱线图反映的数据性质：

(1) 中心位置：中位数 M 所在的位置.

(2)分散程度：全部数据都落在$[Min, Max]$之内，在区间$[Min, Q_1]$，$[Q_1, M]$，$[M, Q_3]$，$[Q_3, Max]$的数据各占1/4，区间较短时，表示落在该区间的数据较集中，反之较为分散；

(3)对称性：若中位数位于箱子的中间位置，则数据分布较为对称。M与Min和Max的距离大小反映了数据分布向左或向右倾斜，且能看出分为尾部的长短。

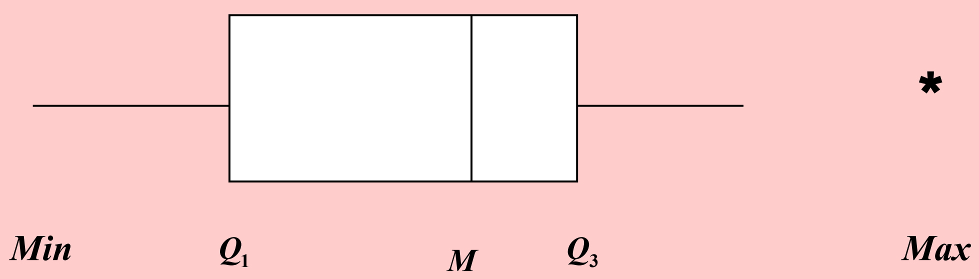
疑似异常值：在数据集中某一个不寻常地大于或小于该数集中地其他数据。

注：疑似异常值会对随后的计算结果产生不适当的影响，有必要对其进行检查和适当的处理。

四分位数间距： $IQR \xrightarrow{\Delta} Q_3 - Q_1$

在实际数据的处理中，若数据小于$Q_1 - 1.5IQR$或大于$Q_3 + 1.5IQR$，就认为它是疑似异常值。

修正箱线图：



§3 统计量与抽样分布

1. 定义：

设 $(X_1,\cdots X_n)$ 为来自总体 X 的一个样本，$g(X_1,\cdots X_n)$ 是 $X_1,\cdots X_n$ 的函数，若 g 中不含任何未知参数；

则称 $g(X_1,\cdots X_n)$ 是一个统计量。

设 $x_1,\cdots,x_n$ 是相应于样本 $(X_1,\cdots X_n)$ 的样本值。

则称 $g(x_1,\cdots x_n)$ 是 $g(X_1,\cdots X_n)$ 的观察值。

注：统计量是随机变量。

返回主目录

例 1

设 $X_1,\cdots X_n$ 为来自总体 $X\sim N(\mu,\sigma^2)$ 的一个样本，其中μ未知,σ^2已知，问下列随机变量中那些是统计量

$$\min(X_1,X_2,\cdots,X_n);\ \frac{X_1+X_n}{2};\ \frac{X_1+\cdots+X_n}{n}-\mu\ ;$$

$$\frac{(X_1+X_n)^2}{\sigma^2};\ \frac{(X_1+\cdots+X_n)-n\mu.}{\sqrt{n}\sigma}.$$

2. 常用的统计量 --- 样本的数字特征

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

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

样本标准差： $S=\sqrt{S^2}=\sqrt{\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}$

样本k阶(原点)矩： $A_k=\dfrac{1}{n}\sum_{i=1}^{n}X_i^{\,k}\quad k=1,2,\cdots$

样本k阶中心矩： $B_k=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k\quad k=1,2,\cdots$

它们的观察值分别为：

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

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

返回主目录

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

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

$$b_k=\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^k,\quad k=1,2\cdots$$

分别称为样本均值、样本方差、样本标准差、样本 k 阶矩、样本 k 阶中心矩。

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。

注意：1)若总体X的k阶(原点)矩

$$E(X^k) \overset{记}{=} \mu_k 存在，\quad k=1,2,\cdots$$

则当$n\to\infty$时，$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k \xrightarrow{P} \mu_k, \quad k=1,2,\cdots$

事实上：$X_1,\cdots,X_n$相互独立，且与总体 X 同分布，

所以 $X_1^k,\cdots,X_n^k$ 相互独立，且与总体 X^k 同分布。

故有 $E(X_1^k)=\cdots=E(X_n^k)=\mu_k \quad k=1,2,\cdots$

由辛钦大数定律知

$$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k \xrightarrow{P} \mu_k, \quad k=1,2,\cdots$$

返回主目录

注意：2）设 $X_1,\cdots X_n$为来自总体X 的一个样本

$$EX=\mu\,,\,DX=\sigma^2,$$

则

$$E\overline{X}=\mu\,,\,D\overline{X}=\frac{\sigma^2}{n},\,ES^2=\sigma^2.$$

证：

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

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

样本方差：$S^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2=\dfrac{1}{n-1}[\sum\limits_{i=1}^{n}X_i^{\,2}-n\overline{X}^2]$

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

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

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

说明： 当总体的方差 σ^2 未知时，可用

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

$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$ 近似代替 σ^2 。

返回主目录

例 2 设总体 $X \sim \exp(\theta)$，$X_1,\cdots,X_n$ 为来自总体 X 的一个样本，则

$$E\overline{X} = E(X) = \theta, \ D\overline{X} = \frac{D(X)}{n} = \frac{\theta^2}{n},$$
$$ES^2 = D(X) = \theta^2.$$

例 3 设总体 $X \sim U(-1,3)$, $X_1,\cdots,X_n$ 为来自总体 X 的一个样本，则

$$E\overline{X} = E(X) = \frac{-1+3}{2}, D\overline{X} = \frac{D(X)}{n} = \frac{1}{n} \times \frac{(3+1)^2}{12},$$

$$ES^2 = D(X) = \frac{(3+1)^2}{12}.$$

返回主目录

例 4 设总体 $X \sim \pi(\lambda)$， $X_1, \cdots, X_n$ 为来自总体 X 的一个样本，则

$$E\overline{X} = E(X) = \lambda, \, D\overline{X} = \frac{D(X)}{n} = \frac{\lambda}{n},$$

$$ES^2 = D(X) = \lambda.$$

经验分布函数

$v_n(x)$表示随机事件$\{X \le x\}$在n次独立重复观测中出现的次数 $\longrightarrow$ **经验频数**

$$v_n(x) \sim B(n, F(x)).$$

经验分布函数：

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

把样本值$x_1, x_2, \cdots, x_n$按值从小到大排序

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

则

$$F_n(x) = \begin{cases} 0 & x < x_{(1)} \\ \dfrac{k}{n} & x_{(k)} \le x < x_{(k+1)} \quad k = 1,2,\cdots,n-1 \\ 1 & x_{(n)} \le x \end{cases}$$

经验分布函数的性质：

(1)　是分布函数

(2)　随机变量(样本函数)：　$E[v_n(x)] = n \cdot F(x)$　$E\left[\dfrac{v_n(x)}{n}\right] = F(x)$

(3)　$\lim\limits_{n\to\infty} P\left\{\left|F_n(x) - F(x)\right| < \varepsilon\right\} = 1 \quad \forall\varepsilon > 0$　依概率收敛

(4)　$F_n(x)$以概率1一致收敛于它的理论分布函数$F(x)$

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

（格里纹科定理）

3. 常用统计量的分布

(1) χ^2 – 分布

设$(X_1,\cdots X_n)$为来自于正态总体$N(0,1)$的样本，则称统计量：

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

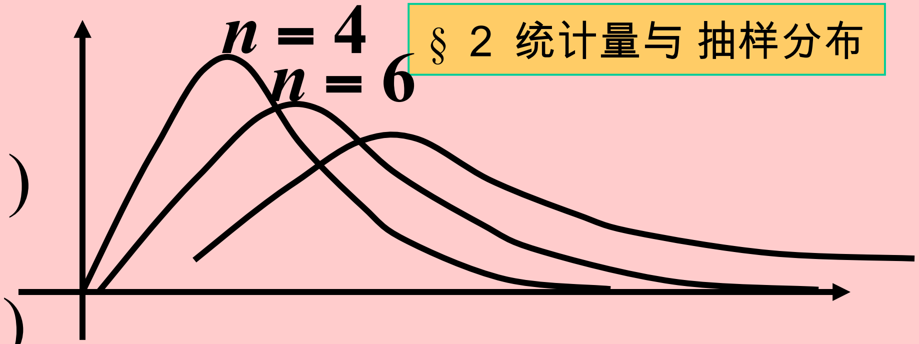
所服从的分布为自由度是n的χ^2分布。

记为 $\chi^2 \sim \chi^2(n)$

χ^2分布的概率密度

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

返回主目录



说明： $\because X_i \sim N(0,1)$,

$\therefore X_i^2 \sim \chi^2(1)$, (p_{51})

即 $X_i^2 \sim \Gamma(\frac{1}{2},2)$ (p_{78})

又由 $X_1^2,\cdots,X_n^2$ 相互独立，及 Γ 分布的可加性得

$$\chi^2(n) = X_1^2 + \cdots + X_n^2 \sim \Gamma(\frac{n}{2},2)$$

特别

$$\chi^2(2) = X_1^2 + X_2^2 \sim \Gamma(1,2) = \exp(2)$$

χ^2分布的性质：

$1^0.\chi_1^2 \sim \chi^2(n_1), \chi_2^2 \sim \chi^2(n_2)$，且$\chi_1^2$，$\chi_2^2$独立，则有

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

$2^0. E\chi^2 = n, \quad D\chi^2 = 2n$

证2^0：

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

$$EX_i = 0, \quad DX_i = 1, \quad X_i \sim N(0,1)$$

$$EX_i^2 = DX_i + (EX_i)^2 = 1,$$

所以 $E\chi^2 = n.$

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

$$DX_i^2 = EX_i^4 - (EX_i^2)^2 = 3 - 1 = 2, \quad i = 1,2,\cdots n$$

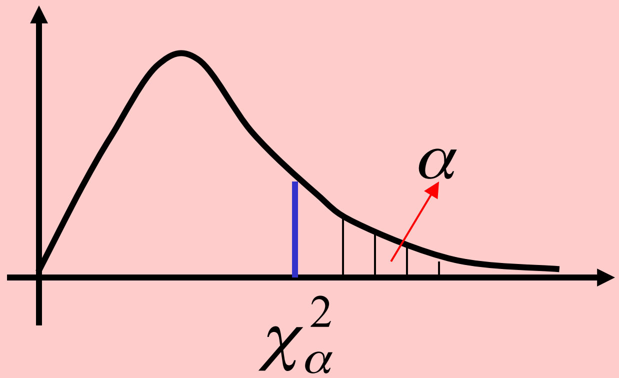
$$EX_i^4 = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} x^4 e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} x^3 d(-e^{-\frac{x^2}{2}})$$

$$= \frac{1}{\sqrt{2\pi}}\{[-x^3 e^{-\frac{x^2}{2}}]_{-\infty}^{+\infty} + 3\times\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}} dx$$

$$= 3EX_i^2 = 3,$$

所以 $\quad D\chi^2 = 2n.$

返回主目录



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$\chi^2(n)$分布的上α分位点。

$$\chi^2_{0.1}(25) = 34.382 \quad , \quad \chi^2_{0.05}(35) = 49.802 \, .$$

当n充分大时，$\chi^2_\alpha(n) \approx \frac{1}{2}(z_\alpha + \sqrt{2n-1})^2$

z_α是标准正态分布的上 α 分位点。

返回主目录

(2) $t-$分布

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

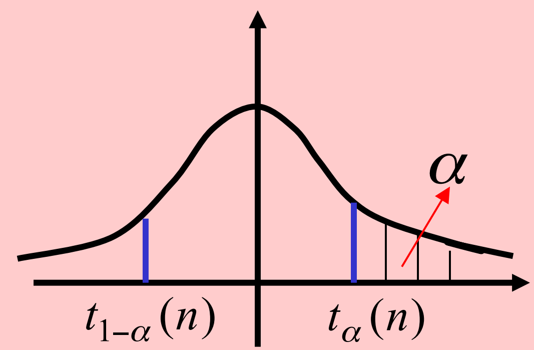
$$t = \frac{X}{\sqrt{Y/n}}$$所服从的分布为自由度是n的 $t-$分布

或称学生氏（ $Student$)分布，记作$t \sim t(n)$.

$t-$分布的概率密度为

$$h(t) = \frac{\Gamma[(n+1)/2]}{\sqrt{\pi n}\,\Gamma(n/2)}\left(1+\frac{t^2}{n}\right)^{-(n+1)/2},$$

$$-\infty < t < \infty$$



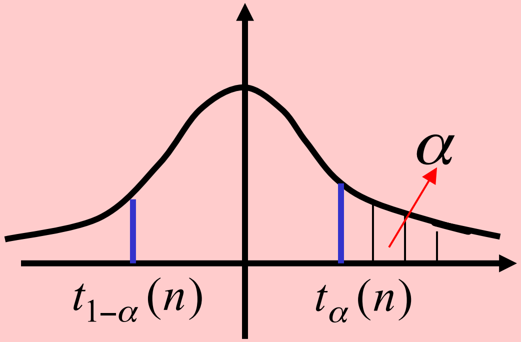
可证 $$\lim_{n\to\infty} h(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

当n很大时，$t-$分布 近似服从标准正态分布 $N(0,1)$.

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：$t_{1-\alpha}(n) = -t_\alpha(n)$

当$n>45$时，$t_\alpha(n) \approx z_\alpha$.

返回主目录

(3) $F-$分布

若 $X\sim\chi^2(n_1),Y\sim\chi^2(n_2)$, X,Y 独立， 则称随机变量

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

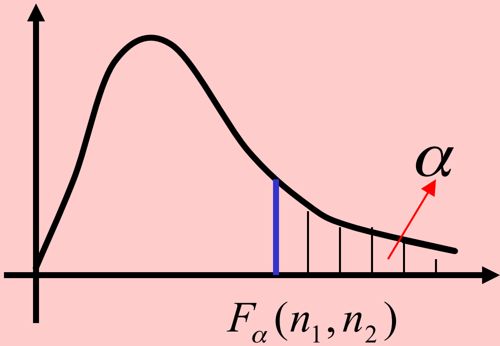
所服从的分布为自由度

是n_1,n_2的 $F-$分布，记作$F\sim F(n_1,n_2)$.

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

对于给定的$\alpha(0<\alpha<1)$，称满足条件：

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



的点$F_\alpha(n_1,n_2)$为F分布的上α分位点。

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

返回主目录

证明：若$F \sim F(n_1, n_2)$

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

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

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

又因为 $1/F \sim F(n_2,n_1)$，所以 $F_\alpha(n_2,n_1) = \frac{1}{F_{1-\alpha}(n_1,n_2)}$

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

例： $F_{0.95}(12,9) = \frac{1}{F_{0.05}(9,12)} = \frac{1}{2.80} = 0.357$

返回主目录

4 、正态总体的样本均值与样本方差的分布：

定理 1. 设$(X_1,\cdots,X_n)$是总体$N(\mu,\sigma^2)$的样本，$\overline{X},S^2$分别是样本均值与样本方差，则有：

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

证 :$X_i\sim N(\mu,\sigma^2),i=1,2,\cdots,n,$ 且它们独立,

所以 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ 服从正态分布，

又知 $E\overline{X}=\mu\,,D\overline{X}=\frac{\sigma^2}{n},$ 故 $\overline{X}\sim N(\mu,\frac{\sigma^2}{n}).$

返回主目录

定理 2. 设$(X_1,\cdots,X_n)$是总体$N(\mu,\sigma^2)$的样本，$\overline{X},S^2$

分别是样本均值与样本方差，则有：

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

(2). $\overline{X}$与S^2独立。

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

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

证明： $$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1), \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

且它们独立。 则由 t- 分布的定义：

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

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

定理 4. 设 $(X_1, X_2, \cdots, X_{n_1})$与$(Y_1, Y_2, \cdots Y_{n_2})$分别是具有两个正态总体 $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$的样本，且它们独立。

设 $$\overline{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i \quad , \quad \overline{Y} = \frac{1}{n_2}\sum_{j=1}^{n_2} Y_j$$

分别是两个样本的均值 。

$$S_1{}^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i - \overline{X})^2 , \quad S_2{}^2 = \frac{1}{n_2 - 1}\sum_{j=1}^{n_2}(Y_j - \overline{Y})^2$$

分别是两个样本的方差 ；

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

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

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

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

且它们独立，

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

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

证 :$\overline{X}-\overline{Y} \sim N(\mu_1-\mu_2, \frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2})$

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

$\frac{(n_1-1)S_1^2}{\sigma^2} \sim \chi^2(n_1-1), \frac{(n_2-1)S_2^2}{\sigma^2} \sim \chi^2(n_2-1)$,且它们独立。

则 $$\frac{(n_1-1)S_1^2}{\sigma^2}+\frac{(n_2-1)S_2^2}{\sigma^2} \sim \chi^2(n_1+n_2-2).$$

返回主目录

由$t-$分布的定义：

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

$$\sim t(n_1+n_2-2)$$

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

例 5 设总体 $X \sim N(0, 0.25)$, $X_1, \cdots, X_5$ 为来自总体 X 的一个样本，求：$P\{\sum_{i=1}^{5} X_i^2 > 0.4025\}$.

解： $\because X_i \sim N(0, 0.25)$ $\quad \therefore \dfrac{X_i - 0}{0.5} = 2X_i \sim N(0,1)$.

$$\therefore 4\sum_{i=1}^{5} X_i^2 \sim \chi^2(5)$$

$$P\{\sum_{i=1}^{5} X_i^2 > 0.4025\} = P\{4\sum_{i=1}^{5} X_i^2 > 1.610\} = 0.90$$

$$\boxed{\chi_\alpha^2(5) = 1.610}$$

返回主目录

例 6 设总体 $X \sim N(12,4)$，$(X_1,\cdots,X_5)$为来自总体X的一个样本，求：(1) $P\{|\overline{X}-12|>1\}$;

(2) $P\{\max(X_1,X_2,X_3,X_4,X_5)>15\}$;

解：(1) $P\{|\overline{X}-12|>1\}=1-P\{|\overline{X}-12|\le 1\}$

$$=1-P\{-\frac{\sqrt{5}}{2}\le\frac{\overline{X}-12}{2/\sqrt{5}}\le\frac{\sqrt{5}}{2}\}$$

$$=2[1-\Phi(\frac{\sqrt{5}}{2})]=0.2628$$

$$\boxed{\overline{X}\sim N(12,\frac{4}{5}).}$$

返回主目录

(2) $P\{\max(X_1,X_2,X_3,X_4,X_5)>15\}$

$=1-\ P\{\max(X_1,X_2,X_3,X_4,X_5)\le 15\}$

$=1-\ P\{X_1\le 15,X_2\le 15,X_3\le 15,X_4\le 15,X_5\le 15\}$

$=1-\ \prod_{i=1}^{5}P\{X_i\le 15\}$

$X_i \sim N(12,4).$

$=1-\prod_{i=1}^{5}P\{\frac{X_i-12}{2}\le\frac{15-12}{2}\}$

$=1-[\Phi(1.5)]^5=0.2923$

例 7 设总体 $X \sim N(\mu, \sigma^2)$，$(X_1, \cdots, X_{16})$ 为来自总体 X 的一个样本，S^2 为样本方差，

求：(1) $P\{S^2/\sigma^2 \le 2.04\}$，(2) $D(S^2)$.

解：(1) 由定理 2 知 $\dfrac{15S^2}{\sigma^2} \sim \chi^2(15)$,

$$\therefore P\{S^2/\sigma^2 \le 2.04\} = P\{15S^2/\sigma^2 \le 15 \times 2.04\}$$

$$= 1 - P\{15S^2/\sigma^2 > 30.615\}$$

$$\approx 1 - 0.01 = 0.99$$

(2) $D(S^2) = D\left(\dfrac{\sigma^2}{15}\chi^2(15)\right) = \dfrac{2\sigma^4}{15}$

返回主目录

例 8设总体 X 与 Y 相互独立，且 $X \sim N(0,1)$，$Y \sim N(0,1)$, $(X_1,\cdots,X_9)$为来自总体 X 的一个样本 $Y_1,\cdots,Y_9$ 为来自总体 Y 的一个样本，则统计量

$$U = \frac{X_1 + \cdots + X_9}{\sqrt{Y_1^2 + \cdots + Y_9^2}}$$

服从 $(A)\ \chi^2(9)$, $(B)\ \chi^2(8)$, $(C)\ t(9)$, $(D)\ t(8)$.

解：

$$\overline{X} = \frac{1}{9}(X_1 + \cdots + X_9) \sim N(0,\frac{1}{9})$$

$$\frac{\overline{X} - 0}{1/3} = 3\overline{X} \sim N(0,1)$$

返回主目录

$$Y_1^2+\cdots+Y_9^2\sim\chi^2(9)$$

$\overline{X}$与$Y_1^2+\cdots+Y_9^2$ 相互独立，

$$U=\frac{X_1+\cdots+X_9}{\sqrt{Y_1^2+\cdots+Y_9^2}}=\frac{3\overline{X}}{\sqrt{(Y_1^2+\cdots+Y_9^2)/9}}\sim t(9)$$

例 9 设总体$X \sim N(\mu,\sigma^2)$，$(X_1,\cdots,X_{10})$为来自总体X的一个样本，$\bar{X} = \frac{1}{9}\sum_{i=1}^{9} X_i$，$S^2 = \frac{1}{8}\sum_{i=1}^{9}(X_i - \bar{X})^2$

证明统计量 $t = \frac{X_{10} - \bar{X}}{S}\sqrt{\frac{9}{10}} \sim t(8)$

解：$\bar{X} = \frac{1}{9}(X_1 + \cdots + X_9) \sim N(\mu,\frac{\sigma^2}{9})$，$X_{10} \sim N(\mu,\sigma^2)$

$$\frac{X_{10} - \bar{X}}{\sqrt{\sigma^2 + \frac{\sigma^2}{9}}} = \frac{X_{10} - \bar{X}}{\sigma}\sqrt{\frac{9}{10}} \sim N(0,1)$$

返回主目录

$$\frac{8S^2}{\sigma^2} \sim \chi^2(8)$$

$$\boxed{S^2 = \frac{1}{8}\sum_{i=1}^{9}(X_i - \overline{X})^2}$$

$X_{10} - \overline{X}$与S^2相互独立

$$t = \frac{X_{10} - \overline{X}}{S}\sqrt{\frac{9}{10}} = \frac{(X_{10} - \overline{X})/\sigma}{\sqrt{S^2/\sigma^2}}\sqrt{\frac{9}{10}} \sim t(8)$$

例 10 设总体 $X \sim N(0,2^2)$， $X_1,\cdots,X_4$ 为来自总体 X 的一个样本，当 a,b 为何值时，统计量

$$Y = a(X_1 - 2X_2)^2 + b(3X_3 - 4X_4)^2$$

服从χ^2分布,其自由度是多少？

解：$E(X_1 - 2X_2) = E(X_1) - 2E(X_2) = 0$

$$D(X_1 - 2X_2) = D(X_1) + 4D(X_2) = 20$$

$$\therefore \quad X_1 - 2X_2 \sim N(0,20)$$

例 10 $\boxed{X \sim N(0,2^2)}$,

$$\boxed{Y = a(X_1 - 2X_2)^2 + b(3X_3 - 4X_4)^2}$$

$$E(3X_3 - 4X_4) = 3E(X_3) - 4E(X_4) = 0$$

$$D(3X_3 - 4X_4) = 9D(X_3) + 16D(X_4) = 100$$

$$\therefore \quad 3X_3 - 4X_4 \sim N(0,100)$$

$$\therefore \frac{X_1 - 2X_2}{\sqrt{20}} \sim N(0,1), \quad \frac{3X_3 - 4X_4}{10} \sim N(0,1)$$

且它们独立,

例 10 $\boxed{\dfrac{X_1 - 2X_2}{\sqrt{20}} \sim N(0,1)\,,\quad \dfrac{3X_3 - 4X_4}{10} \sim N(0,1)}$

$$\left(\frac{X_1 - 2X_2}{\sqrt{20}}\right)^2 + \left(\frac{3X_3 - 4X_4}{10}\right)^2 \sim \chi^2(2)$$

当$a = \dfrac{1}{20}, b = \dfrac{1}{100}$时，

$$Y = a(X_1 - 2X_2)^2 + b(3X_3 - 4X_4)^2$$

服从χ^2分布,其自由度是2.

例 11：设X_1, X_2是来自于总体$N(0.\sigma^2)$的样本，试求

$$Y=\left(\frac{X_1+X_2}{X_1-X_2}\right)^2$$ 的分布.

解：由X_1, X_2是来自于总体$N(0,\sigma^2)$中的样本，可得

$$X_1+X_2 \sim N(0,2\sigma^2) \qquad \frac{X_1+X_2}{\sqrt{2\sigma^2}} \sim N(0,1)$$

$$X_1-X_2 \sim N(0,2\sigma^2) \qquad \frac{X_1-X_2}{\sqrt{2\sigma^2}} \sim N(0,1)$$

由正态分布的性质可知，

(X_1+X_2, X_1-X_2)的联合分布是二元正态 分布。

因为

$$Cov(X_1+X_2,X_1-X_2)=D(X_1)-D(X_2)=0$$

所以X_1+X_2与X_1-X_2相互独立。

由 F 分布的定义

$$Y=\left(\frac{X_1+X_2}{X_1-X_2}\right)^2=\frac{\left(\dfrac{X_1+X_2}{\sqrt{2\sigma^2}}\right)^2}{\left(\dfrac{X_1-X_2}{\sqrt{2\sigma^2}}\right)^2}\sim F(1,1).$$

1 给出了总体、个体、样本和统计量的概念，要掌握样本均值和样本方差的计算及基本性质。

2 引进了χ^2分布、t 分布、 F 分布的定义，会查表计算。

3 掌握正态总体的某些统计量的分布。

作业： P_{147} 3,4,6,7,8,9.

返回主目录