要求小g先求大G
写出定义

，求随机变量 $U=X+Y$ 的概率密度 $g(u)$.

【分析】 $G(u) \triangleq P\{U \le u\} = P\{X+Y \le u\}$

数理统计基础

1 总体与样本

2 数理统计的常用分布

3 正态总体的抽样分布

4 非正态总体的抽样分布

5 分位数

数理统计的分类

- 描述统计学(Descriptive statistics): 对随机现象进行观测、试验，以取得有代表性的观测值.

summarize, organize and simplify data.

- 推断统计学(Inferential statistics): 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.

generalizations about population parameters based on sample statistics.

推断统计学包括参数估计、假设检验、回归分析、方差分析等内容.

总体、个体、样本

总体(population)——研究对象的全体构成的集合。

个体(individual)——组成总体的每一个成员.

通常我们研究的都是一些数量指标，其数值随着个体的不同而不同.

例： 某城市在职职工的年收入情况，

某中学高二学生的身高与体重情况.

研究的所有成员的数量指标全体称为总体。

把每一个成员的数量指标称为个体。

可以用随机变量X(或随机向量$X=(X_1,\ldots,X_k)$)描述总体（k维总体）.

随机变量X的分布函数称为**总体的分布函数**(理论分布).

抽样

抽样(sampling)是从总体中通过一定的方法选择一部分个体。

样本(sample)：从总体X中抽取的待考查个体组成的集合。
样本中个体的数量n称为样本容量. 容量为n的样本常记为$(X_1, X_2, \ldots, X_n)$. 这里的 X_i 可看作一个随机变量.

样本值：样本一旦经过考查，得到的是n个具体的数$(x_1, x_2, \ldots, x_n)$，称为样本的一次观察值，简称样本值.

样本空间：样本所有可能取值的集合.

抽样的目的是基于样本在某种程度上能代表总体，根据给定的样本对总体进行统计推断。为了使抽取的样本能很好地反映总体的信息，必须考虑抽样方法.

最常用的一种抽样方法叫作“简单随机抽样”，要求样本$X_1, X_2, \ldots, X_n$满足下面两点:

① 代表性：样本$X_1, X_2, \ldots, X_n$中的每个X_i与所考察的总体具有相同的分布.

② 独立性：样本$X_1, X_2, \ldots, X_n$是相互独立的.

以后若无特别说明，样本都是指简单随机样本.

- 有限总体
  - 采用放回抽样就能得到简单随机样本.
  - 当总体容量很大的时候,放回抽样有时候不方便。当总体容量比较大时(总体中个体总数/样本容量≥10),通常将不放回抽样所得到的样本近似当作简单随机样本来处理.
- 无限总体，一般采取不放回抽样.

简单随机样本的性质

1. 若总体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

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

2. 若总体X的密度函数为$f(x)$，则样本的联合密度函数为

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

3. 设总体X的均值μ，方差σ^2存在， 则$E(X_i) = \mu$，$D(X_i) = \sigma^2$

例 设 $(X_1, X_2, \cdots, X_n)$ 是取自正态总体 $N(\mu; \sigma^2)$ 的样本, 这个样本的联合概率密度函数为

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

$$= \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right],$$

$$-\infty < x_1, x_2, \cdots, x_n < +\infty.$$

样本空间是R^n

例 某厂要检查一批产品的质量, 每件产品可区分为合格品与不合格品, 用数字“0”表示合格品, 用数字“1”表示不合格品,这时, 总体 X 服从 0-1 分布:

$$P\{X=x\}=\begin{cases} p, x=1 \\ 1-p, x=0 \end{cases}$$

其中 p 表示一件产品为不合格品的概率. 上式可改写为

$$P\{X=x\}=p^x(1-p)^{1-x},\ x\in\{0,1\}.$$

若 $(X_1, X_2, \cdots, X_n)$ 是取自这个总体 X 的一个样本, 则有

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

$$=p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i},\ x_i\in\{0,1\}, i=1,2,\ldots,n$$

样本的经验分布函数

总体分布函数 $F(x)=P(X\le x)$

设总体 X 的分布函数为 $F(x)$ (未知), 其样本观测值为 $x_1,x_2,\cdots,x_n$, 将它们从小到大排列成 $x_{(1)}\leqslant x_{(2)}\leqslant\cdots\leqslant x_{(n)}$. 设其中互不相同的共 l 个, 分别为 $x_{(1)}^*<x_{(2)}^*<\cdots<x_{(l)}^*$, 其个数分别为 $n_1,n_2,\cdots,n_l,\sum_{i=1}^{l}n_i=n$.

$$F_n(x)=\begin{cases}0, & x<x_{(1)}^*,\\ \dfrac{n_1+n_2+\cdots+n_k}{n}, & x_{(k)}^*\leqslant x<x_{(k+1)}^*,\ k=1,2,\cdots,l-1,\\ 1, & x\geqslant x_{(l)}^*,\end{cases}$$

称为该样本的**经验分布函数**(Empirical cdf).

经验分布函数 $F_n(x)$ 具有下列性质：

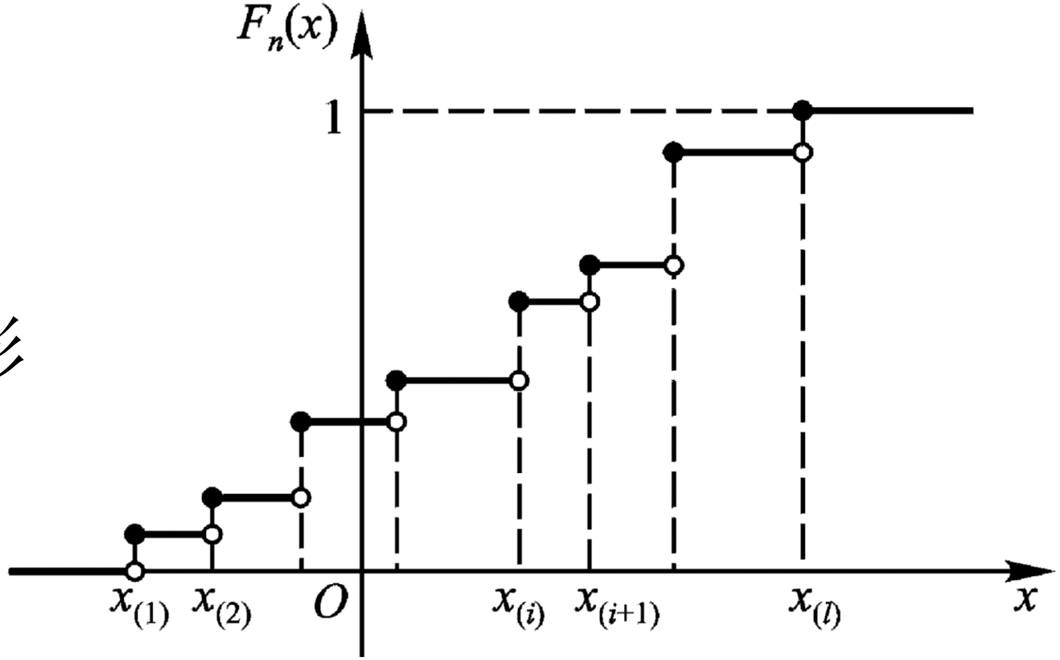
(1) $0 \le F_n(x) \le 1$

(2) $F_n(x)$ 是非减函数；

(3) $F_n(-\infty) = 0, \quad F_n(+\infty) = 1$

(4) $F_n(x)$ 在每个观测值 $x_{(i)}$ 处是右连续的，点 $x_{(i)}$ 是 $F_n(x)$ 的跳跃间断点，$F_n(x)$ 在该点的跃度就等于频率 f_i

经验分布函数 $F_n(x)$ 的图形



例 从某个总体 X 中抽取容量为10的一个样本，测得样本观测值为 $-4,0,1,-2,5,3,3,5,6,5$. 试求该样本的经验分布函数.

解 将样本观测值从小到大的顺序排列为

$$-4 \quad -2 \quad 0 \quad 1 \quad 3 \quad 3 \quad 5 \quad 5 \quad 5 \quad 6$$

其经验分布函数为

$$F_n(x)=\begin{cases}0, & x<-4,\\ 0.1, & -4\leqslant x<-2,\\ 0.2, & -2\leqslant x<0,\\ 0.3, & 0\leqslant x<1,\\ 0.4, & 1\leqslant x<3,\\ 0.6, & 3\leqslant x<5,\\ 0.9, & 5\leqslant x<6,\\ 1, & x\geqslant 6.\end{cases}$$

格利文科定理（Glivenko 1933）

经验分布 $F_n(x)$ 是样本 $(X_1, X_2, \cdots, X_n)$ 中不大于 x 的个体数与 n 之比, 即在 n 次重复独立试验中事件 $\{X \leqslant x\}$ 发生的频率;

总体 X 的分布函数 $F(x)$ 是事件 $\{X \leqslant x\}$ 发生的概率;

当 n 充分大时, $F_n(x)$ 是否趋向于 $F(x)$ 呢?

Glivenko定理（格林文科定理）

经验分布 $F_n(x)$ 以概率为 1 地一致收敛于 $F(x)$, 即

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

定理表明, $F_n(x)$ 与 $F(x)$ 之差的绝对值在 $(-\infty, +\infty)$ 上的上确界当 $n \to +\infty$ 时, 极限为 0 的概率是 1. 因此, 只要 n 充分大,$F_n(x)$ 是 $F(x)$ 的很好的近似, 而$F_n(x)$ 可由样本观测值得到的. 这是**用样本推断总体**的理论基础.

概率论中的三种收敛形式

- 分布收敛(convergence in distribution)

$$\lim_{n\to\infty} P\left(\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \le x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt = \Phi(x)$$

$$\boxed{\sum_{k=1}^{n} X_k \xrightarrow{d} N(n\mu, n\sigma^2)}$$

- 依概率收敛(convergence in probability)

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| < \varepsilon\right) = 1$$

$$\boxed{\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{p} \mu}$$

- 几乎处处收敛(convergence almost surely，以概率为1地收敛)

$$P\left\{\lim_{n\to\infty} X_n = X\right\} = 1$$

$$\boxed{X_n \xrightarrow{a.s.} X}$$

统计量

A *parameter* defines a characteristic of the population.
A *statistic* defines a characteristic of the sample.

样本来自于总体，含有总体的信息，但较为分散. 为了进行统计推断，需要把分散的信息集中起来，针对不同的研究目的，构造不同的样本函数，这类函数称为统计量.

设$X_1, X_2, \ldots, X_n$是来自总体X的一个样本，$T(X_1, X_2, \ldots, X_n)$是一个实值函数，且不含任何未知参数，则称函数$T(X_1, X_2, \ldots, X_n)$为一个统计量(statistic).

若$(x_1, \ldots, x_n)$是样本观测值，则称$T(x_1, \ldots, x_n)$为统计量$T(X_1, X_2, \ldots, X_n)$的一个观测值.

说明：统计量 $T(X_1, X_2, \ldots, X_n)$ 是随机变量;
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.

常用的统计量（1）--样本矩

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

(1) $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ **样本均值(sample mean)**

(2) $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ **样本方差(sample variance)**

(3) $S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$ **样本标准差** (Sample Standard deviation)

(4) $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ 为样本的k阶原点矩 $A_1 = \overline{X}$

(5) $B_k = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^k$ 为样本的k阶中心矩 $B_2 = \dfrac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \triangleq S_n^2$

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

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

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

Bessel's Correction

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

对于二维总体(X,Y)，常用的统计量有

（1）样本协方差

$$S_{XY}^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})$$

（2）样本相关系数

$$\rho_{XY}=\frac{S_{XY}^2}{S_X S_Y}=\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\cdot\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}}$$

当样本取得观测值$x_1,x_2,\dots,x_n$后，代入可得到这些统计量的观测值.

例　设总体X的均值μ，方差σ^2存在，$(X_1,\cdots,X_n)$是取自总体X的样本，$\bar{X},S^2$为样本均值和样本方差；求：$E(\bar{X}),D(\bar{X}),E(S^2)$

解：$$E(\bar{X})=E(\frac{1}{n}\sum_{i=1}^{n}X_i)=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=\mu,$$

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

对于正态分布 $X\sim N(\mu,\sigma^2)$，期望 $E(X^2)$ 可以通过以下公式计算：

首先，使用期望的性质：

$$E(X^2)=\mathrm{Var}(X)+(E(X))^2$$

对于正态分布 $X\sim N(\mu,\sigma^2)$，我们知道：

- $E(X)=\mu$
- $\mathrm{Var}(X)=\sigma^2$

因此：

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

所以，对于正态分布 $X\sim N(\mu,\sigma^2)$，有：

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

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

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

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

常用的统计量（2）--次序统计量

设 $(X_1, X_2, \cdots, X_n)$ 是取自总体 X 的样本,

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

（1）最小顺序统计量 $\boldsymbol{X_{(1)} = \min\{X_1, X_2, \ldots, X_n\}}$

（2）最大顺序统计量 $\boldsymbol{X_{(n)} = \max\{X_1, X_2, \ldots, X_n\}}$.

（3）**样本极差** $R = X_{(n)} - X_{(1)}$

（4）**样本中位数**

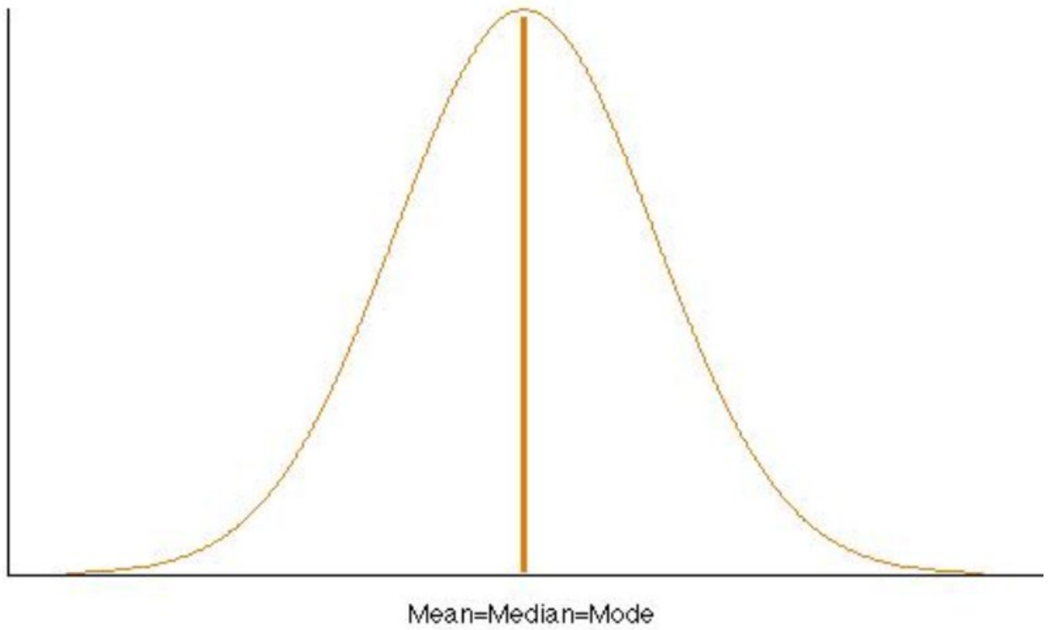
$$\operatorname{med}(X_1, X_2, \cdots, X_n) = \begin{cases} X_{\left(\frac{n+1}{2}\right)}, & n \text{ 为奇数}, \\ \dfrac{1}{2}\left(X_{\left(\frac{n}{2}\right)} + X_{\left(1+\frac{n}{2}\right)}\right), & n \text{ 为偶数} \end{cases}$$

样本极差与样本中位数分别反映了数据在数轴上分布的离散性特征与位置特征.

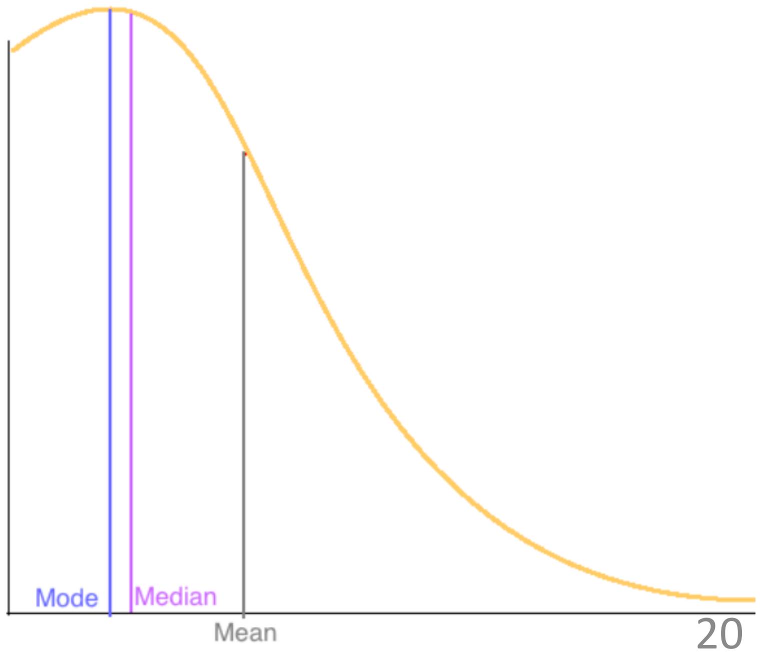
3M

- Mean(样本均值)
- Median(样本中位数): 从小到大排列的数列中位于中间位置的那个数，如果数列个数为奇数位，则中位数位于(n+1)/2；如果为偶数位，则为n/2与n/2+1的两个数的平均值。
- Mode(众数)：出现次数最多的值

正态分布：mean = median = mode



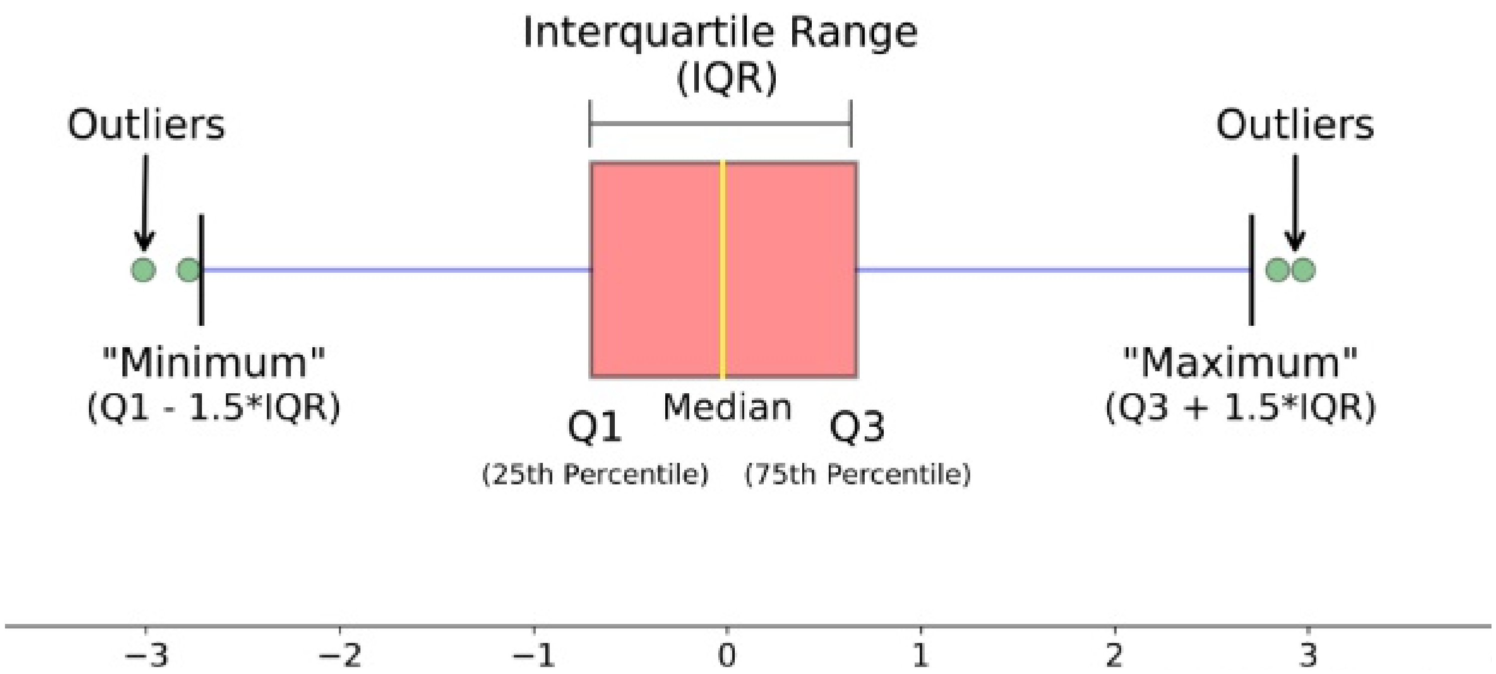
非正态分布



箱线图(boxplot)

四分位数：将从小到大排列的数列平均分为4段，最小的四分位数称为**下四分位数**(q1)，最大的称为**上四分位数**(q3)，中间的称为**中位数**(q2, median)。

- 四分位距IQR=Q3-Q1, interquartile range
- 上限=Q3+1.5IQR，Q3向上延伸1.5倍IQR的数据点.
- 下限=Q1-1.5IQR，Q1向下延伸1.5倍IQR的数据点.
- **离群值**（显示为绿色圆圈）

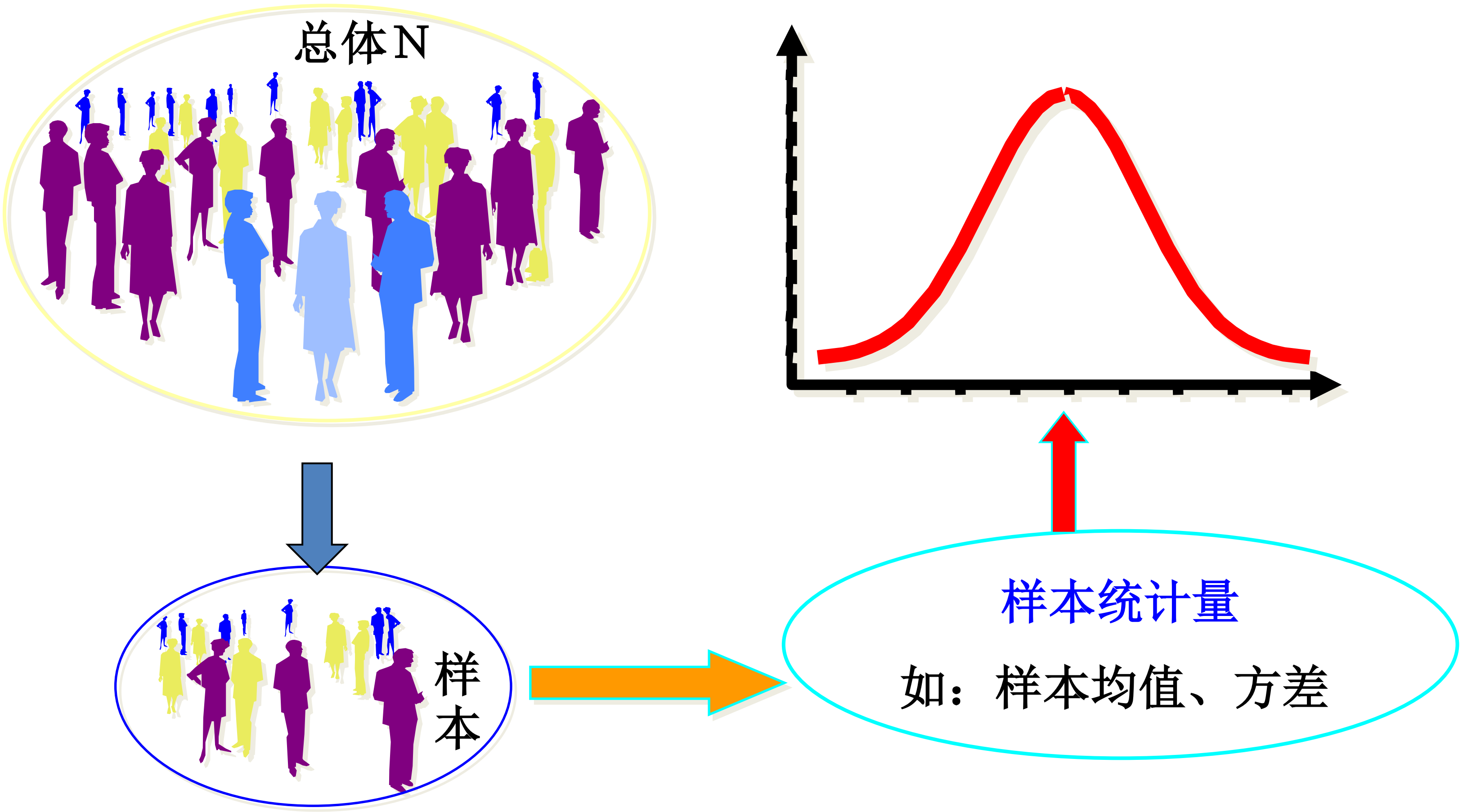


2　数理统计的常用分布

- 正态分布
- $\boldsymbol{\chi^2}$分布
- $\boldsymbol{t}$分布
- $\boldsymbol{F}$分布

抽样分布

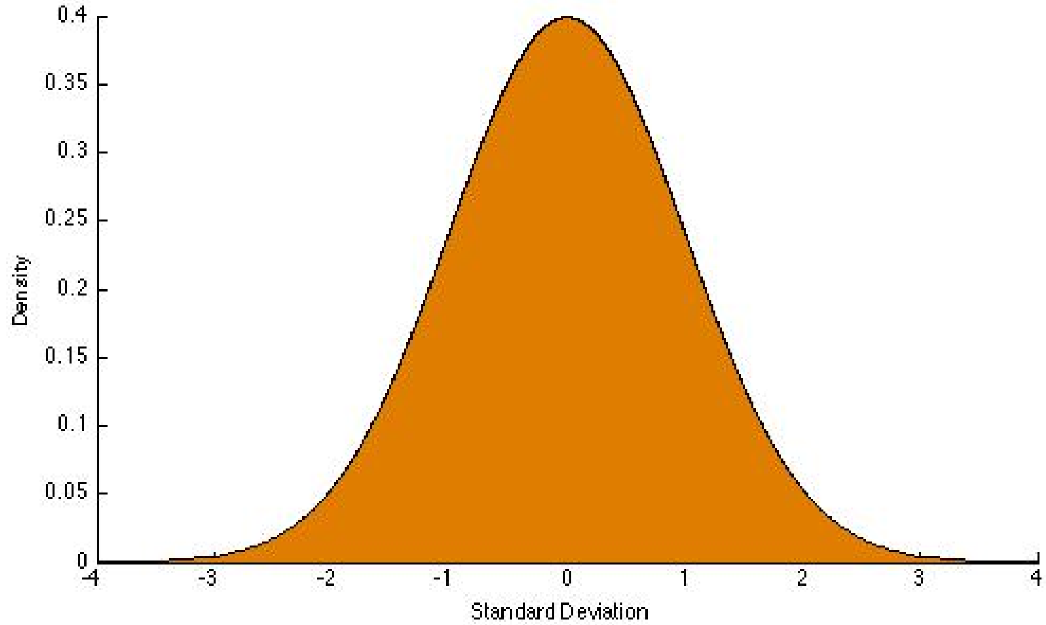
统计量依赖于样本，统计量是随机变量，统计量的分布叫做***抽样分布***(sampling distribution)．



与正态分布有关的常用抽样分布

（1）标准正态分布

$$X \sim N(0,1)$$



密度函数为

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < +\infty$$

$$\boxed{X \sim N(\mu, \sigma^2) \Longleftrightarrow \frac{X-\mu}{\sigma} \sim N(0,1)}$$

(2) χ^2分布

设 $X_1, X_2, \cdots, X_n$相互独立，都服从正态分布$N(0,1)$，

则随机变量：

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

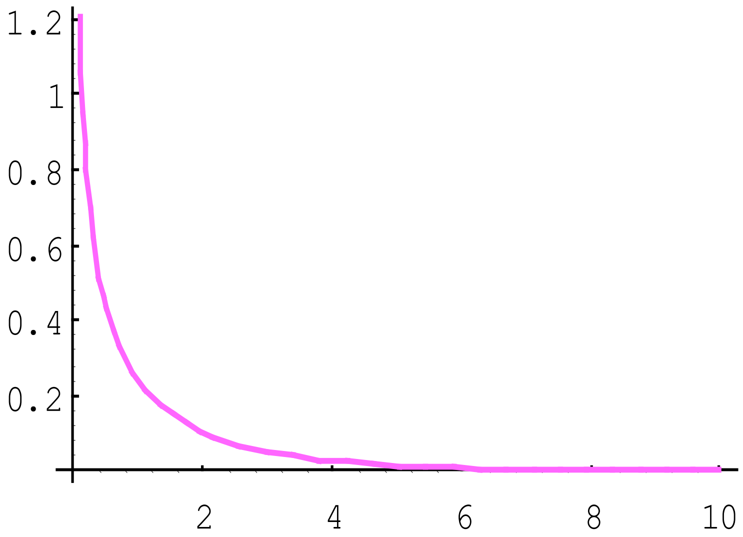
服从自由度为 n 的χ^2分布. 记为 $\chi^2 \sim \chi^2(n)$

degrees of freedom informally corresponds to the number of "independent pieces of information"

海尔墨特(Hermert)和现代统计学奠基人之一的卡尔·皮尔逊(Karl Pearson)分别于1875年和1900年推导出来。

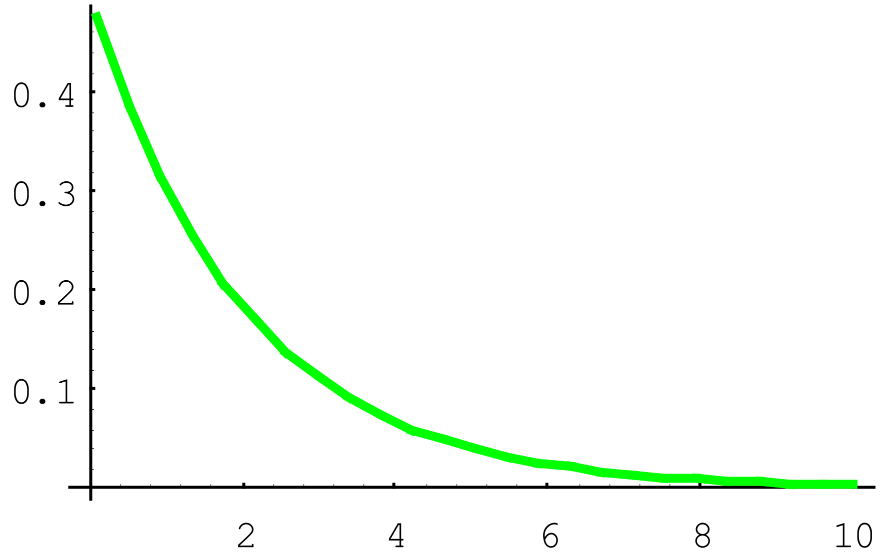
$n=1$ 时,其密度函数为

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



$n=2$ 时,其密度函数为

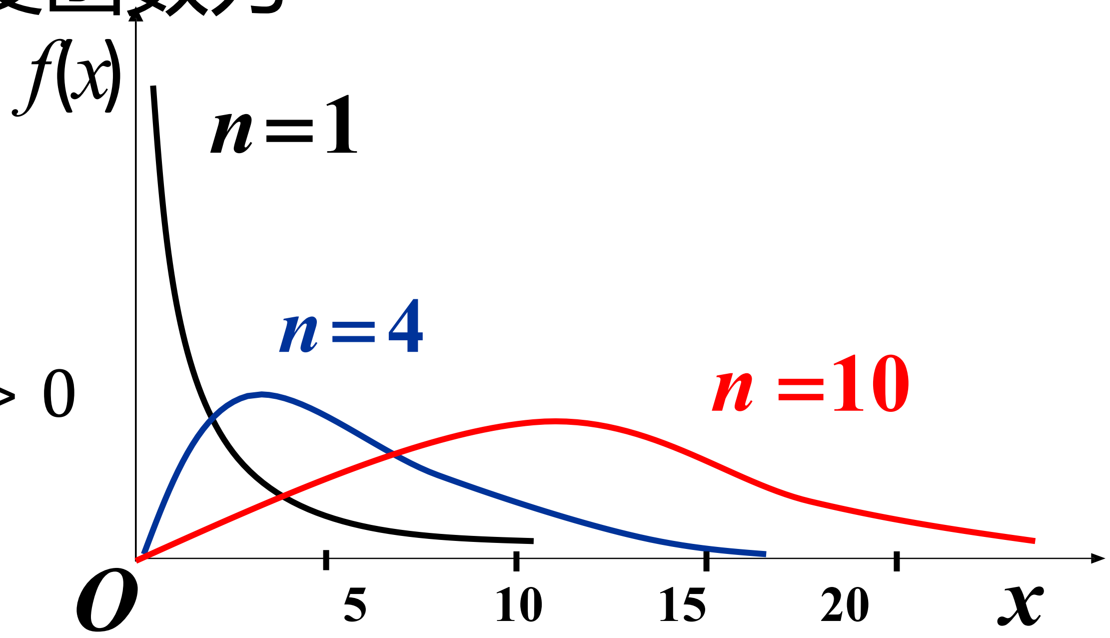
$$f(x)=\begin{cases}\dfrac{1}{2}e^{-\frac{x}{2}}, & x>0\\ 0, & x\le 0\end{cases}$$



即参数为1/2的指数分布

自由度为 n 的$\chi^2(n)$的密度函数为

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



χ^2分布的概率密度函数

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

$\Gamma(x+1)=x\Gamma(x)$, $\Gamma(0)=\Gamma(1)=1, \Gamma(n+1)=n! \quad (n\in N)$

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

证 因为 $X_1, X_2, \cdots, X_n$ 相互独立, 且都服从 $N(0;1)$, 所以 $\boldsymbol{X} = [X_1, X_2, \cdots, X_n]^{\mathrm{T}}$ 的概率密度为

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

当 $x \geqslant 0, \Delta x > 0$ 时, 随机事件 $\{x < \chi^2 \leqslant x + \Delta x\}$ 的概率为

$$P\{x < \chi^2 \leqslant x + \Delta x\} = (2\pi)^{-\frac{n}{2}} \iint_G \cdots \int \exp\left(-\frac{1}{2}\sum_{i=1}^{n} x_i^2\right) \mathrm{d}x_1\, \mathrm{d}x_2 \cdots \mathrm{d}x_n,$$

其中 G 是外半径为 $\sqrt{x + \Delta x}$ 、内半径为 $\sqrt{x}$ （$x = \sum_{i=1}^{n} x_i^2$ 是一个常数）的 n 维球壳. 记 G 的体积为 ΔV, 则有

$$(2\pi)^{-\frac{n}{2}} \mathrm{e}^{-\frac{1}{2}(x+\Delta x)} \Delta V \leqslant P\{x < \chi^2 \leqslant x + \Delta x\} \leqslant (2\pi)^{-\frac{n}{2}} \mathrm{e}^{-\frac{1}{2}x} \Delta V.$$

由于半径为r的n维球体的体积是 Cr^n, 这里的C是常数, 故$\Delta V = C\left[(x+\Delta x)^{\frac{n}{2}} - x^{\frac{n}{2}}\right]$, 从而

$$\lim_{\Delta x\to 0}\frac{\Delta V}{\Delta x} = C\lim_{\Delta x\to 0}\frac{1}{\Delta x}\left[(x+\Delta x)^{\frac{n}{2}} - x^{\frac{n}{2}}\right] = C\frac{n}{2}x^{\frac{n}{2}-1}.$$

因而，$\lim\limits_{\Delta x\to 0}\frac{1}{\Delta x}P\{x<\chi^2\leqslant x+\Delta x\} = kx^{\frac{n}{2}-1}\mathrm{e}^{-\frac{x}{2}}$，其中 $k=(2\pi)^{-\frac{n}{2}}\frac{n}{2}C$ 是常数.

当 $x>0$ 时, χ^2 的概率密度是 $kx^{\frac{n}{2}-1}\mathrm{e}^{-\frac{x}{2}}$.

因 $\chi^2\geqslant 0$, 故当 $x\leqslant 0$ 时, χ^2 的概率密度显然为0 .

由概率密度的性质知,

$$1=\int_{-\infty}^{+\infty} f(x)\mathrm{d}x = \int_0^{+\infty} kx^{\frac{n}{2}-1}\mathrm{e}^{-\frac{x}{2}}\,\mathrm{d}x = k2^{\frac{n}{2}}\int_0^{+\infty} u^{\frac{n}{2}-1}\mathrm{e}^{-u}\,\mathrm{d}u = k2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right),$$

即 $k=\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}$. 于是，$f(x)=\begin{cases}\frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}x^{\frac{n}{2}-1}\mathrm{e}^{-\frac{x}{2}}, & x>0,\\ 0, & x\leqslant 0.\end{cases}$

χ^2分布的性质：

性质1　变量值始终为正；

性质2　$E(\chi^2(n)) = n$，$D(\chi^2(n)) = 2n$；

性质3　χ^2分布的可加性

设X~ $\chi^2(n_1)$, Y~ $\chi^2(n_2)$, 并且相互独立，则X+Y ~$\chi^2(n_1+n_2)$

若 $X_i \sim \chi^2(n_i)(i=1,2,\cdots,k)$, 且 $X_1, X_2, \cdots, X_k$ 相互独立,
则 $\sum_{i=1}^{k} X_i \sim \chi^2\left(\sum_{i=1}^{k} n_i\right)$

性质4　χ^2分布的极限分布是正态分布。

设$\mathbf{X} \sim \chi^2(n)$，则对任意$x$，有

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

即　$\chi^2(n) \overset{\text{近似}}{\sim} N(n, 2n)$.

χ^2分布的数学期望和方差

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

证明 因为 $X_i \sim N(0, 1)$,所以 $E({X_i}^2) = D(X_i) = 1$

$$E\left(X_i^4\right) = \int_{-\infty}^{+\infty} x^4 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \mathrm{d}x = 3$$

$$D({X_i}^2) = E({X_i}^4) - [E\left({X_i}^2\right)]^2 = 3 - 1 = 2，\quad i = 1, 2, \cdots, n.$$

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

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

(3) t 分布 (Student 分布)

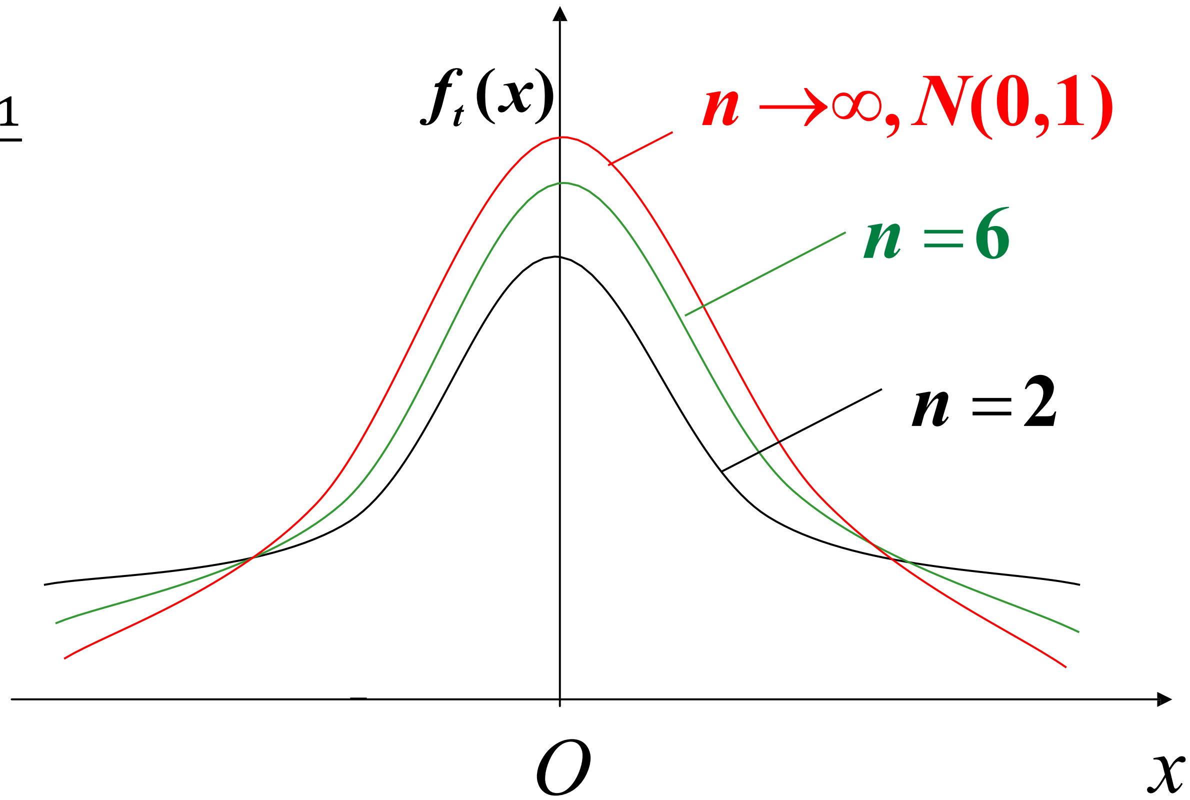
设$X \sim N(0,1)$，$Y \sim \chi^2(n)$，X, Y独立，则称随机变量 $\mathbf{T} = \dfrac{X}{\sqrt{Y/n}}$

为服从自由度是$\boldsymbol{n}$的$\boldsymbol{t}$分布，记作$T \sim t(n)$.

其密度函数为

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

$$-\infty < x < \infty$$



t分布的密度函数： 低峰、厚尾

t分布的性质

性质1 密度函数$f(x,n)$是偶函数，且 $\lim\limits_{n\to\infty} f(x,n)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}=\varphi(x)$

即t分布的极限分布是标准正态分布.

当n充分大(n>=30)时，t 分布近似$N(0,1)$分布.

对于较小的n，t分布与$N(0,1)$分布相差很大.

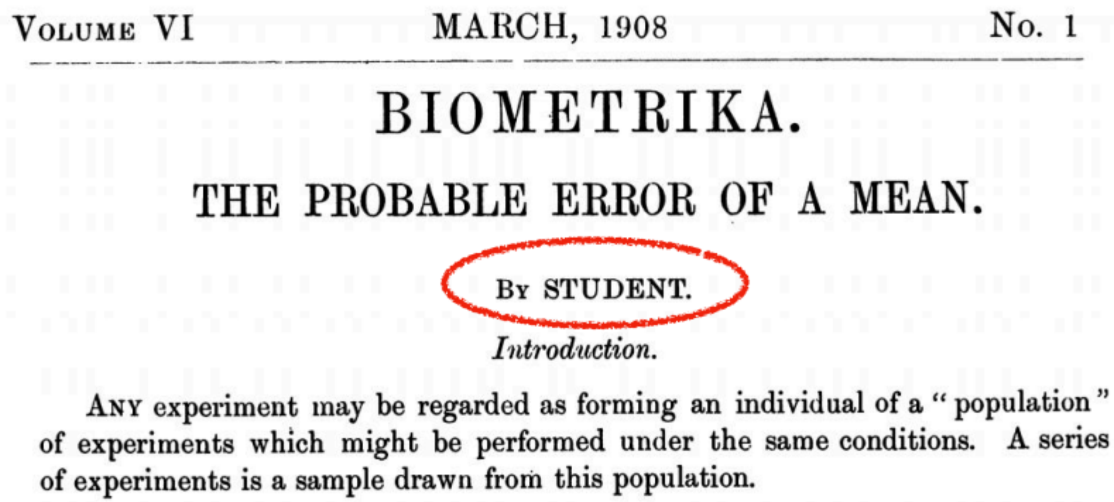
性质2 设$T\sim t(n)$，则

当$n=1$时， $E(T)$不存在, $t(1)$是标准柯西分布,

当$n\geq 2$时， $E(T)=0$,

当$n\geq 3$时， $D(T)=\dfrac{n}{n-2}$.

“student”：
William Sealy Gosset(1876-1937)



VOLUME VI MARCH, 1908 No. 1

BIOMETRIKA.

THE PROBABLE ERROR OF A MEAN.

BY STUDENT.

Introduction.

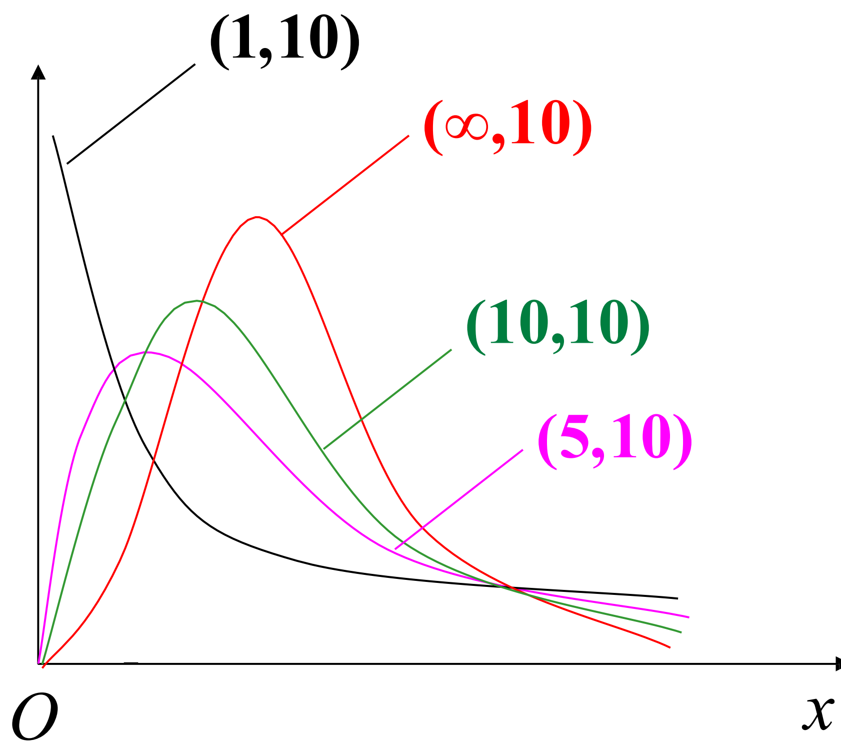
ANY experiment may be regarded as forming an individual of a “population” of experiments which might be performed under the same conditions. A series of experiments is a sample drawn from this population.

(4) F 分布

若 $X\sim\chi^2(m)$, $Y\sim\chi^2(n)$, X,Y 独立，则称随机变量 $\mathbf{F}=\frac{X/m}{Y/n}$ 为服从**自由度是(m,n)的F分布**，记作$F\sim F(m,n)$. m称为第一自由度，n称为第二自由度。

概率密度函数$f(x)$为

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



F分布是以英国统计学家费歇（R.A. Fisher，1890－1962）命名的.

F分布的性质：

(1) 若 $F \sim F(n_1, n_2)$, 则 $\dfrac{1}{F} \sim F(n_2, n_1)$

简证：$F \sim F(n_1, n_2)$, $\exists U \sim \chi^2(n_1)$, $V \sim \chi^2(n_2)$, 使$F = \dfrac{U/n_1}{V/n_2}$

$$\Rightarrow \frac{1}{F} = \frac{V/n_2}{U/n_1} \sim F(n_2, n_1)$$

(2) 若 $t \sim t(n)$, 则 $t^2 \sim F(1, n)$

简证：$t \sim t(n) \Rightarrow \exists X \sim N(0,1)$, $Y \sim \chi^2(n)$, 使$t = \dfrac{X}{\sqrt{Y/n}}$

$$t^2 = \frac{X^2}{Y/n}, \quad X^2 \sim \chi^2(1), \ Y \sim \chi^2(n), \ F\text{分布定义}$$

(3) $E(F) = \dfrac{n}{n-2}\ (n > 2)$, $D(F) = \dfrac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$, $(n > 4)$.

3 正态总体的抽样分布

理论上，当总体的分布已知时，统计量的分布可以通过随机变量函数的分布求出。但由于涉及复杂的运算，往往得不到统计量分布的解析表达式。

正态总体是最常见的总体，在正态总体条件下，某些统计量的分布有比较简单的形式（四大分布）。

3.1 单个正态总体的抽样分布定理

设 $X_1, X_2, \cdots, X_n$ 是来自正态总体 $X \sim N\left(\mu, \sigma^2\right)$ 的一组样本, $\bar{X}, S^2$ 分别是样本均值和样本方差, 则

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

与中心极限定理的结论对比

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

(3) $\bar{X}$ 与 S^2 相互独立;

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

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

证明

(1) 若$X_1, X_2, \cdots, X_n \sim N(\mu_i, \sigma_i^2)$ 且独立, 则

$$\sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

若$X_1, X_2, \cdots, X_n$ i.i.d， $X_i \sim N(\mu, \sigma^2)$，则

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

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

(2) 思路： $X_i \xrightarrow{\text{标准化}} Y_i \xrightarrow{Y_i-\bar{Y}\ \text{独立化}} Z_i$

令 $Y_i = \frac{X_i-\mu}{\sigma}(i=1,2,\cdots,n)$，则 $Y_1,Y_2,\cdots,Y_n$ 相互独立，且具有相同的分布 $N(0,1)$，

$\sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(n)$。 $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{\bar{X}-\mu}{\sigma}$

取一个n 阶**正交矩阵A**：

$$\begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix} = A\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{pmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{2\cdot 1}} & -\frac{1}{\sqrt{2\cdot 1}} & 0 & \cdots & 0 \\ \frac{1}{\sqrt{3\cdot 2}} & \frac{1}{\sqrt{3\cdot 2}} & -\frac{2}{\sqrt{3\cdot 2}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & -\frac{n-1}{\sqrt{n(n-1)}} \end{pmatrix}\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$

$$Z_1 = \frac{1}{\sqrt{n}}Y_1 + \frac{1}{\sqrt{n}}Y_2 + \cdots + \frac{1}{\sqrt{n}}Y_n = \sqrt{n}\bar{Y} = \sqrt{n}\frac{\bar{X}-\mu}{\sigma} \sim N(0,1)$$

$$Z_i = a_{i1}Y_1 + a_{i2}Y_2 + \cdots + a_{in}Y_n \sim N(0,1), \quad i = 2,3,\cdots,n$$

由 $\boldsymbol{Y} \sim N_n(\mathbf{0}, \mathbf{E}_n)$ 知, $E(\mathbf{Z}) = E(\mathbf{AY}) = \mathbf{A}E(\mathbf{Y}) = \mathbf{0}$,

$$D(\mathbf{Z}) = \mathrm{Cov}(\mathbf{Z}, \mathbf{Z}) = \mathrm{Cov}(\mathbf{AY}, \mathbf{AY}) = \mathbf{A}\mathrm{Cov}(\mathbf{Y}, \mathbf{Y})\mathbf{A}^{\mathrm{T}} = \mathbf{AE}_n\mathbf{A}^{\mathrm{T}} = \mathbf{E}_n,$$

故 $Z_1, Z_2, \cdots, Z_n$两两不相关(相互独立), 而且 $Z_i \sim N(0,1)(i = 1,2,\cdots,n)$.

根据正交变换不改变向量的长度, 有 $\sum_{i=1}^{n} Y_i^2 = \sum_{i=1}^{n} Z_i^2$.

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

$$= \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = \sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2$$

$$= \sum_{i=1}^{n} Z_i^2 - Z_1^2 = \sum_{i=2}^{n} Z_i^2 \sim \chi^2(n-1).$$

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

(3) 由于 $Z_1, Z_2, \cdots, Z_n$ 相互独立, 故 Z_1 与 $\sum_{i=2}^{n} Z_i^2$ 相互独立,

$\bar{X} = \frac{\sigma Z_1}{\sqrt{n}} + \mu$只依赖于$Z_1$, $S^2 = \frac{\sigma^2 \sum_{i=2}^{n} Z_i^2}{(n-1)}$ 只依赖于$Z_2, \cdots, Z_n$,

所以 $\bar{X}$ 与 S^2 相互独立.

(4) 由于 $U = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$, $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$, 而且两者相互独立, 根据 t 分布的定义得

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

例　设 $(X_1,\cdots,X_n)$是取自正态总体$X\sim N\left(\mu,\sigma^2\right)$的样本，$S^2$为样本方差，求: $E\left(S^2\right),D(S^2)$

解

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

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

$E\left(S^2\right)=\sigma^2\ ,D\left(S^2\right)=\dfrac{2\sigma^4}{n-1}$

3.2 两个正态总体的抽样分布定理

设$X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$， 且$X$与$Y$独立,

$X_1, X_2, \ldots, X_m$ 是取自X的样本， $Y_1, Y_2, \ldots, Y_n$ 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_X^2和S_Y^2分别是这两个样本的样本方差,则有

(1) $$\frac{S_X^2 / \sigma_1^2}{S_Y^2 / \sigma_2^2} \sim F(m-1, n-1)$$

其中 $S_W^2 = \dfrac{(m-1)S_X^2 + (n-1)S_Y^2}{m+n-2}$

(2) 当$\sigma_1^2 = \sigma_2^2 = \sigma^2$时， $$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_W \sqrt{1/m + 1/n}} \sim t(m+n-2).$$

pooled t-test

(3) 当$m = n$ 时 $$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S/\sqrt{n}} \sim t(n-1)$$

其中 $S^2 = S_X^2 + S_Y^2 - 2S_{XY}^2$

paired t-test

(4) $$\frac{[(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)]}{\sqrt{\dfrac{\sigma_1^2}{m} + \dfrac{\sigma_2^2}{n}}} \sim N(0,1)$$

证明

(1) 由定理1的结论 (2) 知,

$$\frac{(m-1)S_X^2}{\sigma_1^2} \sim \chi^2(m-1),\ \frac{(n-1)S_Y^2}{\sigma_2^2} \sim \chi^2(n-1),$$

而 S_X^2 只依赖于 $X_1, X_2, \cdots, X_m$, S_Y^2 只依赖于 $Y_1, Y_2, \cdots, Y_n$, 从而 S_X^2 与 S_Y^2 相互独立, 由 F 分布的定义知,

$$\boxed{\frac{\frac{(m-1)S_X^2}{\sigma_1^2}/(m-1)}{\frac{(n-1)S_Y^2}{\sigma_2^2}/(n-1)} = \frac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2} \sim F(m-1, n-1).}$$

对比

$$\boxed{\frac{\frac{1}{m}\frac{\sum_{i=1}^{m}(X_i-\mu_1)^2}{\sigma_1^2}}{\frac{1}{n}\frac{\sum_{j=1}^{n}(Y_j-\mu_2)^2}{\sigma_2^2}} = \frac{A}{\frac{\sigma_1^2}{\sigma_2^2}} \sim F(m,n), \quad \text{A}=\frac{\frac{1}{m}\sum_{i=1}^{m}(X_i-\mu_1)^2}{\frac{1}{n}\sum_{j=1}^{n}\left(Y_j-\mu_2\right)^2}}$$

(2) 因为 $\bar{X}-\bar{Y} \sim N\left(\mu_1-\mu_2, \dfrac{\sigma^2}{m}+\dfrac{\sigma^2}{n}\right)$

所以 $U = \dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{m}+\frac{1}{n}}} \sim N(0,1)$ 结论4

由 $\dfrac{(m-1)S_1^2}{\sigma^2} \sim \chi^2(m-1)$, $\dfrac{(n-1)S_2^2}{\sigma^2} \sim \chi^2(n-1)$,

且它们相互独立，故由 χ^2 分布的可加性知

$$V = \frac{(m-1)S_1^2}{\sigma^2} + \frac{(n-1)S_2^2}{\sigma^2} \sim \chi^2(m+n-2)$$

由于 U 与 V 相互独立，按 t 分布的定义．

$$\frac{U}{\sqrt{V/(m+n-2)}} = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{m}+\frac{1}{n}}} \sim t(m+n-2).$$

其中 $S_w = \sqrt{\dfrac{(m-1)S_1^2+(n-1)S_2^2}{(m+n-2)}}$

(3) 当m=n时，令$Z_i=X_i-Y_i$ $\quad Z_i \sim N\left(\mu_1-\mu_2,\ {\sigma_1}^2+{\sigma_2}^2\right)$

且相互独立，看作是来自总体$Z \sim N\left(\mu_1-\mu_2,\ {\sigma_1}^2+{\sigma_2}^2\right)$的样本

$$\frac{\bar{Z}-(\mu_1-\mu_2)}{S_Z/\sqrt{n}} \sim t(n-1)$$

$$\overline{\mathrm{Z}}=\bar{X}-\bar{Y}$$

$$S_Z^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(Z_i-\overline{\mathrm{Z}}\right)^2=\frac{1}{n-1}\sum_{i=1}^{n}\left[(X_i-Y_i)-(\bar{X}-\bar{Y})\right]^2=S^2=\boldsymbol{{S_X}^2+{S_Y}^2-2{S_{XY}}^2}$$

$$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{S/\sqrt{n}} \sim t(n-1)$$

有两个独立总体 $X \sim N(2\mu, \sigma^2), Y \sim N(\mu, \sigma^2)$,

$X_1, \ldots, X_4$ 与 $Y_1, \ldots, Y_9$ 分别是来自总体X与Y的简单随机样本,

$\bar{X}, \bar{Y}$ 分别是样本均值, S_1^2, S_2^2 分别是样本方差,

则 $\frac{S_1^2}{S_2^2} \sim F($ [填空1] , [填空2] $)$

作答

4 非正态总体的抽样分布

对于非正态总体，要求出统计量的分布是比较困难的，即使有时理论上可以求出精确的分布，但其形式复杂而难以应用，这时一般利用中心极限定理推出 $n \to +\infty$ 下的极限分布. 当样本容量很大时，应用有关**统计量的极限分布**作为其近似分布.

定理 设总体X的一阶矩、二阶矩存在: $E(X) = \mu, 0 < D(X) = \sigma^2 < +\infty$, $(X_1, X_2, \cdots, X_n)$ 是取自总体 X 的样本，则当 n 充分大时，

(1) 样本均值 $\bar{X}_n$ 近似地服从 $N\left(\mu; \frac{\sigma^2}{n}\right)$,

(2) 样本方差S^2依概率收敛于 σ^2,

(3) $\sqrt{n}\frac{\bar{X}_n - \mu}{S}$ 近似地服从 $N(0; 1)$,

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

5　分位数(quantile)

设随机变量 X 的分布函数为 $F(x)$, p是一百分数(percentile),

$0 < p < 1$, 若 x_p 使

$$P\{X \leqslant x_p\} = F(x_p) = p,$$

则称 x_p 为此分布的 ***p* 分位数**.

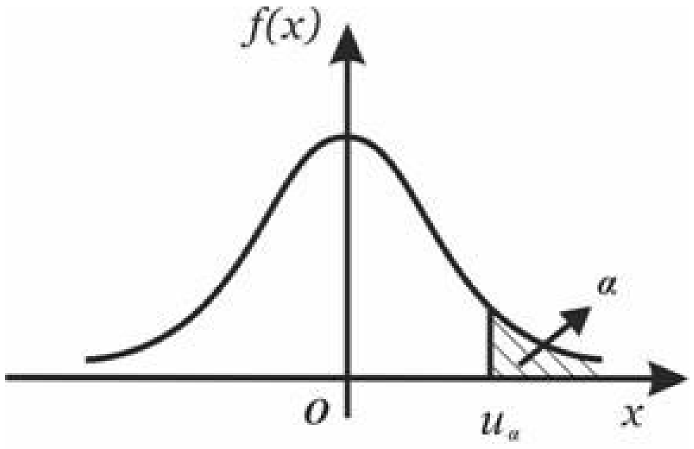
设X的分布函数为$F(x)$，若实数x_α满足

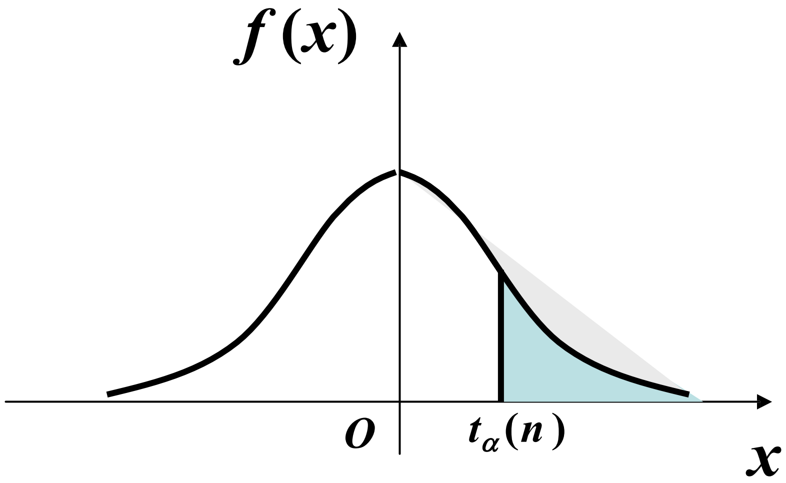
$$P(X > x_\alpha) = 1 - F(x_\alpha) = \alpha,$$

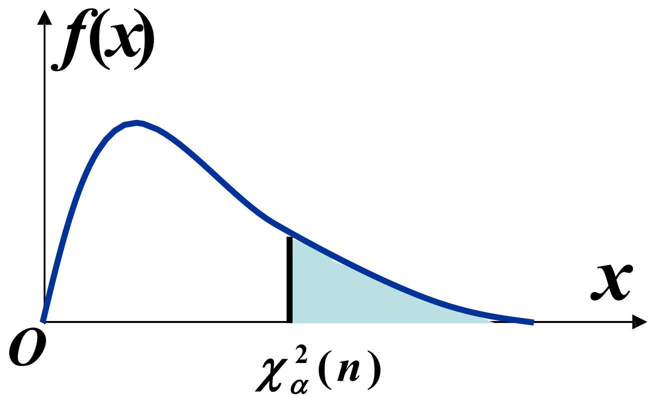
则称x_α为随机变量X的上α分位数.　其中$0 < \alpha < 1$.

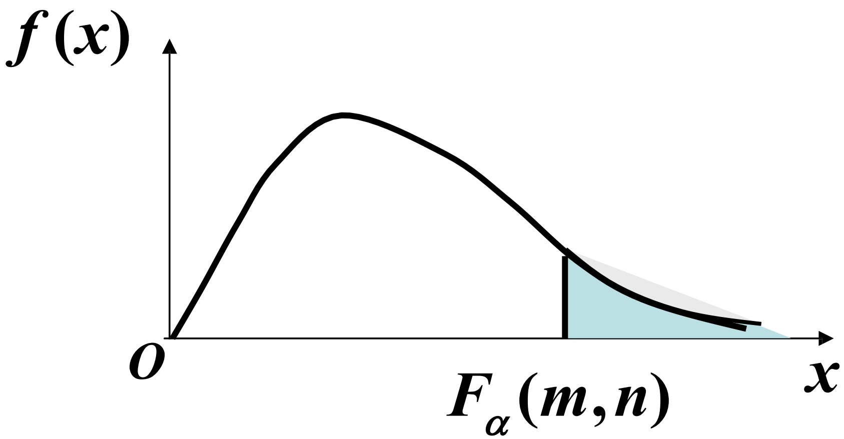
$N(0,1)$，$\chi^2(n)$，$t(n)$，$F(m, n)$的上α分位数分别记为：

u_α，$\chi^2_\alpha(n)$，$t_\alpha(n)$，$F_\alpha(m,n)$







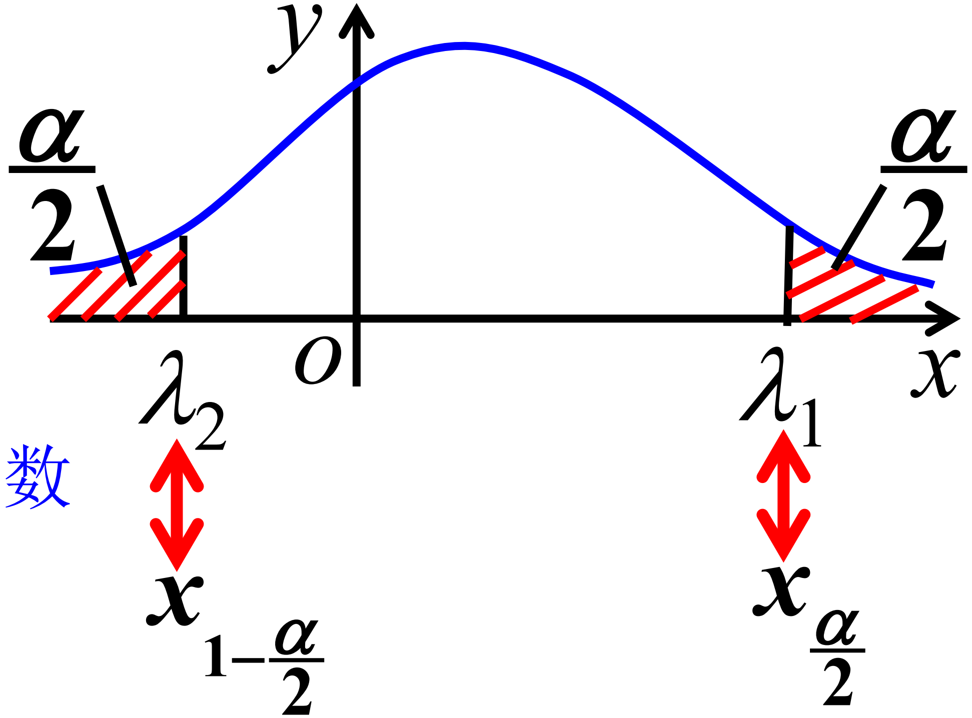


双侧α 分位数

若存在数λ_1、λ_2，使

$$P\{X\geq\lambda_1\}=P\{X\leq\lambda_2\}=\frac{\alpha}{2}$$

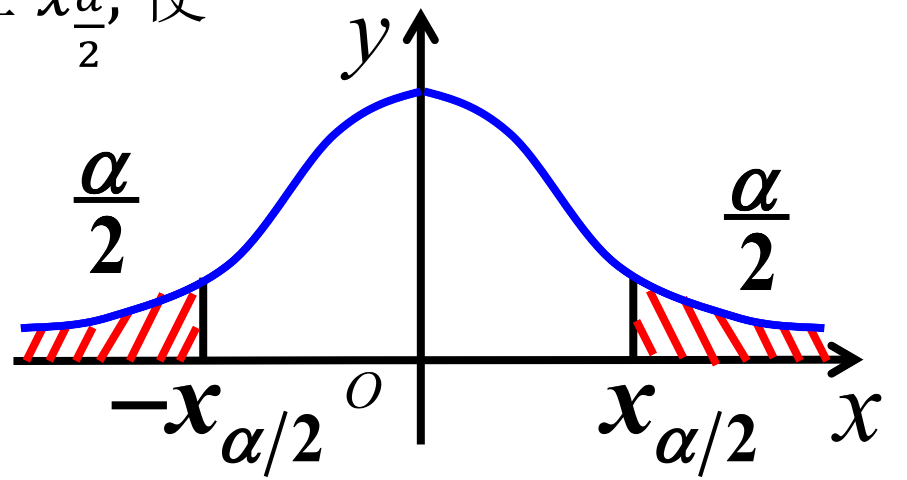
则称λ_1、λ_2为X分布的双侧α分位数



当X的分布关于y轴对称时，若存在 $x_{\frac{\alpha}{2}}$, 使

$$P\{|X|\geq x_{\frac{\alpha}{2}}\}=\alpha,$$

则称$x_{\frac{\alpha}{2}}$ 为X分布的双侧α分位数

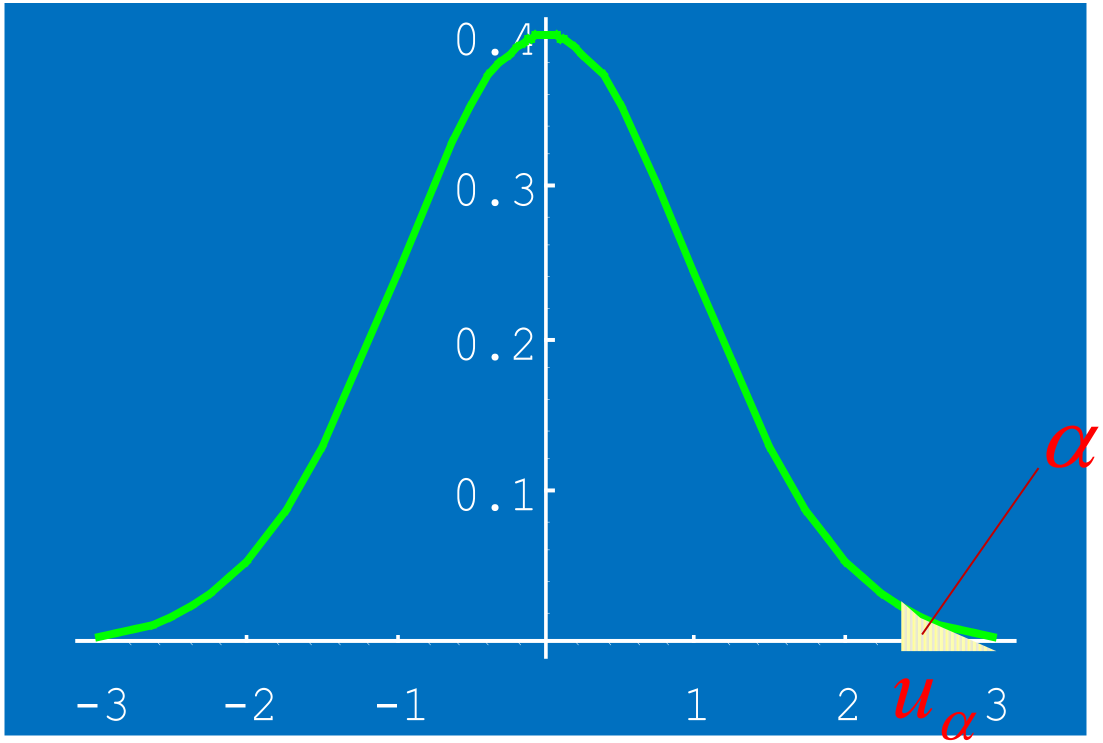


标准正态分布的上 α 分位数(点) u_α

设 $X \sim N(0,1)$, $0<\alpha<1$，称满足

$$P(X > u_\alpha) = \alpha$$

的点 u_α 为X的上α分位数(点)



$$\Phi(u_\alpha) = \int_{-\infty}^{u_\alpha} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{u^2}{2}} \,\mathrm{d}u$$

$$= 1 - \alpha$$

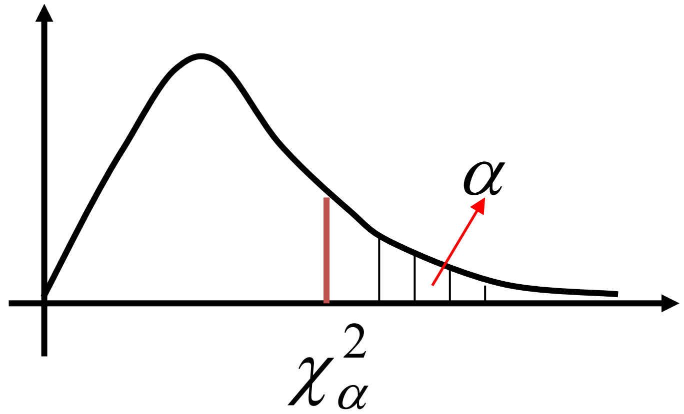
常用数据

$$u_{0.05} = 1.645$$
$$u_{0.025} = 1.96$$
$$u_{0.01} = 2.33$$

对于给定的$\alpha(0<\alpha<1)$，称满足条件：

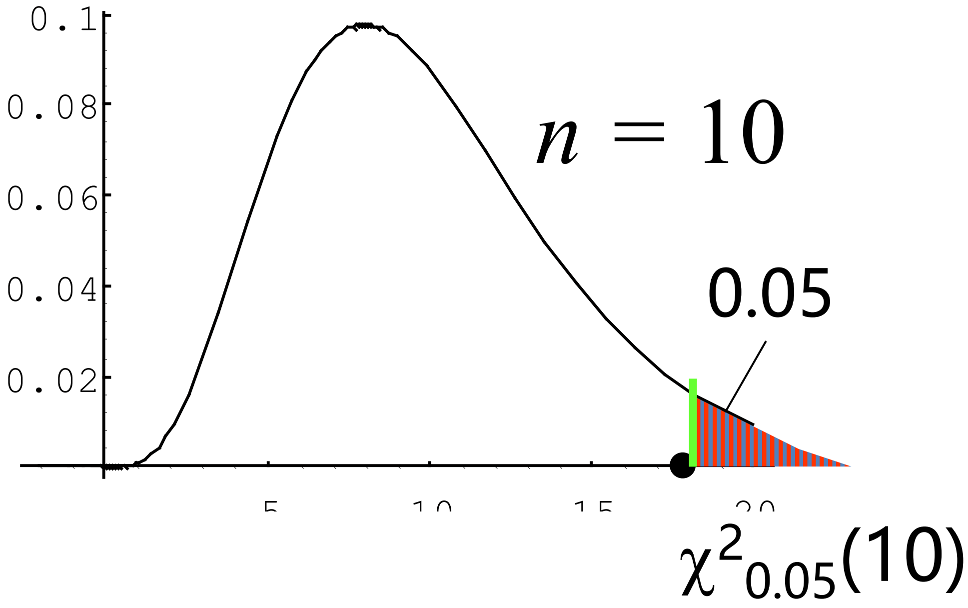
$$P\{\chi^2>\chi^2_\alpha(n)\}=\alpha$$

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



例如

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

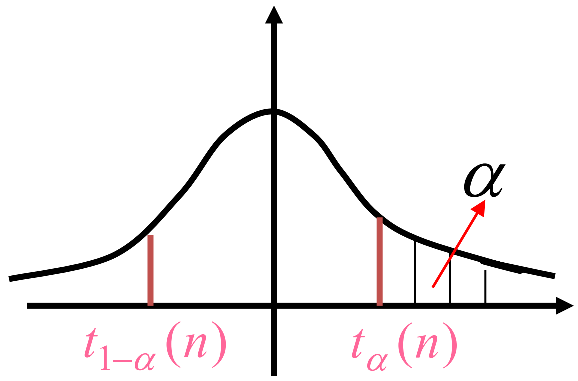


Excel function： chisq.inv(1 - α, k)

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。



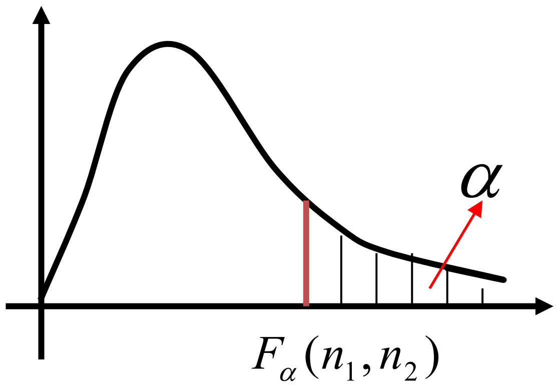
由概率密度的对称性知：$t_{1-\alpha}(n)=-t_{\alpha}(n)$

Excel function t.inv(1 - *α*, *k*)

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。



Excel function f.inv(1 - α, n, m)

分位数的性质：

（1） $u_{1-\alpha}=-u_{\alpha}$， $t_{1-\alpha}(n)=-t_{\alpha}(n)$. X的概率密度关于y轴对称

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

（2）当n充分大（$n>40$即可），有

$$\chi_{\alpha}^{2}(n)\approx\frac{1}{2}(u_{\alpha}+\sqrt{2n-1})^{2}.$$

（3） $F_{1-\alpha}(m,n)=1/F_{\alpha}(n,m)$

F 分布的上α分位点具有性质：$F_{1-\alpha}(n_1,\ n_2) = \dfrac{1}{F_\alpha(n_2, n_1)}$

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

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

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

又因为 $\dfrac{1}{F} \sim F(n_2,\ n_1)$, 所以 $P\left\{\dfrac{1}{F} > F_\alpha(n_2,\ n_1)\right\} = \alpha$,

比较后得， $\dfrac{1}{F_{1-\alpha}(n_1, n_2)} = F_\alpha(n_2,\ n_1)$, 即 $F_{1-\alpha}(n_1,\ n_2) = \dfrac{1}{F_\alpha(n_2, n_1)}$

用来求分布表中未列出的一些上 α 分位点.

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

填空题 1分

设置

根据随机变量的上α分位数的定义，当α增大时，

上α分位数 x_α 将如何变化？

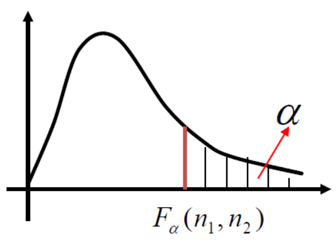
(A) 增大　　(B) 不变　　(C) 减小

[填空1] (填ABC之一)

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。



作答