§2 常用统计分布

在20世纪以前，统计学所处理的数据一般都是大量的、自然采集的，所用的方法以拉普拉斯中心极限定理为依据，总是归结到正态。

到了19世纪末期，数据与正态拟合不好的情况也日渐为人们所注意，进入20世纪之后，人工试验条件下所得数据的统计分析问题，日渐被人们所重视。由于试验数据量有限，那种依赖于近似正态分布的传统方法开始招致质疑，这促使人们研究这种情况下正确的统计方法问题。

统计学三大分布$\chi 2$分布(**Pearson**), *t*分布(**Cosset**)及*F*分布(**Fisher**).

统计学三大剑客(英国):

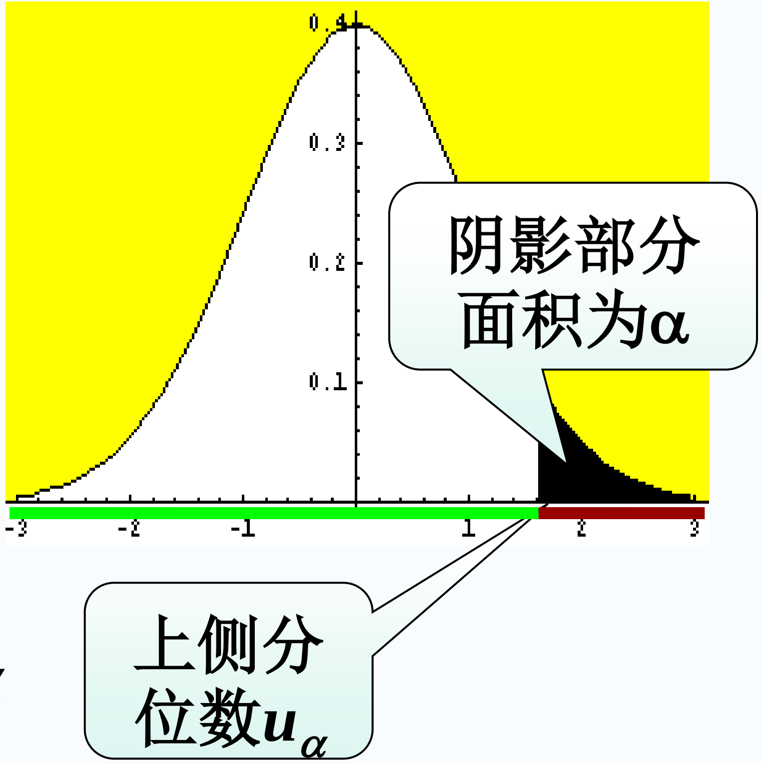
皮尔逊(**Pearson**)、戈塞特(**Cosset**)和费希尔(**Fisher**).

§2 常用统计分布

一、四个常用统计分布

1. 标准正态分布 $X \sim N(0,1)$

$$f(x)=\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}},x\in R$$



上侧分位数u_α (0< α<1):

$$P\{X>u_\alpha\}=\int_{u_\alpha}^{+\infty}f(x)dx=\alpha$$

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

对于标准正态分布有: $\boxed{\Phi(u_\alpha)=1-\alpha}$

$\boxed{\Phi(u_\alpha)=1-\alpha}$ 查表：如 $\alpha=0.025$ 时，$u_\alpha=$ ？

$\Phi(u_{0.025})=1-0.025=0.975 \Rightarrow u_{0.025}=1.96$

2. χ^2（卡方）分布(K.Pearson:英国最高水平统计学家)

最早发现这个分布的其实是物理学家麦克斯韦，他在推导空气分子的运动速度的分布时，发现分子速度在三个坐标轴上的分量是正态分布，而分子运动速度的平方v^2符合自由度为3的χ2分布。

麦克斯韦虽然发现了这个分布，但是真正把他完善并推广的是皮尔逊。

就是在数据挖掘中经常出现的，就是那个皮尔逊相关系数的那个人。

2. χ^2（卡方）分布(K.Pearson:英国最高水平统计学家)

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

自由度为n的含义

其中：$\Gamma(\alpha)=\int_0^{+\infty}x^{\alpha-1}e^{-x}dx,\ \ \alpha>0$为Gama函数.

称随机变量X服从自由度为n 的χ^2分布, 记 $\chi^2\sim\chi^2(n)$

注: 在分布曲线和数据拟合优度检验中χ2分布可是一个利器。

而且皮尔逊的这个工作被认为是假设检验的开山之作。

$\chi^2(n)$自由度为n的含义

自由度是由线性代数中借用的术语

- 定义1. 若存在一组不全为零的常数C_1、C_2、…、C_n 使得 $C_1X_1+C_2X_2+\ldots+C_nX_n=0$，则称变量 X_1、X_2、…、X_n 之间存在一个约束条件.
- 定义2. 若存在 k 个约束条件

$$C_{i1}X_1+C_{i2}X_2+\ldots+C_{in}X_n=0, \quad i=1, 2, \ldots, k$$

其中系数矩阵 $(C_{ij})_{kn}$ 的秩为 k, 且对于任何 m ($k\le m$) 个约束：

$\chi^2(n)$自由度为n的含义

$$C_{i1}'X_1+C_{i2}'X_2+\ldots+C_{in}'X_n=0,\ i=1,2,\ldots,k$$

矩阵$(C_{ij}')_{kn}$的秩总不大于k，则称变量X_1、X_2、…、X_n之间存在 k 个独立的线性约束条件.

易知，X_1、X_2、…、X_n 中只有 $(n-k)$ 个独立变量.

- **定义3.** 若在平方和$\sum X_i^2$中，X_1、X_2、…、X_n之间存在着 k 个独立的线性约束条件，则称平方和$\sum X_i^2$的自由度为 $n-k$ (即其中独立变量为$n-k$ 个).

2. χ^2（卡方）分布(K.Pearson:英国最高水平统计学家)

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

自由度为n的含义

其中：$\Gamma(\alpha)=\int_0^{+\infty}x^{\alpha-1}e^{-x}dx$，$\alpha>0$为Gama函数.

称随机变量X服从自由度为n 的χ^2分布, 记 $\chi^2\sim\chi^2(n)$

Γ函数的主要性质:

$$\Gamma(1)=1,\Gamma(\frac{1}{2})=\sqrt{\pi},\Gamma(\alpha)=(\alpha-1)\Gamma(\alpha-1),\text{ 其中}\alpha>1.$$

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

称随机变量X服从自由度为n的χ^2分布.

Γ函数:$\Gamma(1)=1,\Gamma(\dfrac{1}{2})=\sqrt{\pi},\Gamma(\alpha)=(\alpha-1)\Gamma(\alpha-1)$, 其中$\alpha>1$.

例: 若$X\sim N(0,1)$, 则 $Y=X^2$ 的概率密度为:

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

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

$Y=X^2$ 服从$\chi^2(1)$分布.

$$f(x)=\begin{cases}\dfrac{1}{2\Gamma(\frac{n}{2})}(\dfrac{x}{2})^{\frac{n}{2}-1}e^{-\frac{x}{2}}, & x>0\\ 0, & x\le 0\end{cases}$$，称随机变量X服从自由度为n的χ^2分布.

Γ函数:$\Gamma(1)=1,\Gamma(\frac{1}{2})=\sqrt{\pi},\Gamma(\alpha)=(\alpha-1)\Gamma(\alpha-1)$, 其中$\alpha>1$.

例：若$X_1\sim N(0, 1)$, $X_2\sim N(0,1)$, X_1, X_2相互独立,

则$Y=X_1^2+X_2^2$有

$$f_Y(y)=\begin{cases}\frac{1}{2}e^{-\frac{y}{2}}, & y>0;\\ 0, & y\le 0.\end{cases}=\begin{cases}\frac{1}{2\Gamma(1)}(\frac{y}{2})^{\frac{2}{2}-1}e^{-\frac{y}{2}}, & y>0;\\ 0, & y\le 0.\end{cases}$$

即$Y=X_1^2+X_2^2\sim\chi^2(2)$分布.

定理6.2.1 (结构定理)

设 $X_1, X_2, \ldots, X_n$ 相互独立,

且都服从标准正态分布, 则

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

即随机变量 χ^2 服从自由度为 n 的卡方分布.

例 统计量的分布 (之一)

例：统计量的分布(之一)

设$X_1, X_2, \ldots, X_n$是来自正态总体$X \sim N(\mu,\sigma^2)$的容量为 n 的样本, 求下列统计量的概率分布:

1. $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$　　　2. $Y = \dfrac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$

解： 1. $X_i \sim N(\mu,\sigma^2)$

$$\Rightarrow \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$$

(正态分布的可加性)

故 $\dfrac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu, \dfrac{\sigma^2}{n})$（正态分布的线性性）

设$X_1, X_2, \ldots, X_n$是来自正态总体$X \sim N(\mu,\sigma^2)$的容量为 n 的样本，求下列统计量的概率分布:

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

$\because \quad \dfrac{X_i-\mu}{\sigma}\sim N(0,1)$ 且 $\dfrac{X_i-\mu}{\sigma}$相互独立

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

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

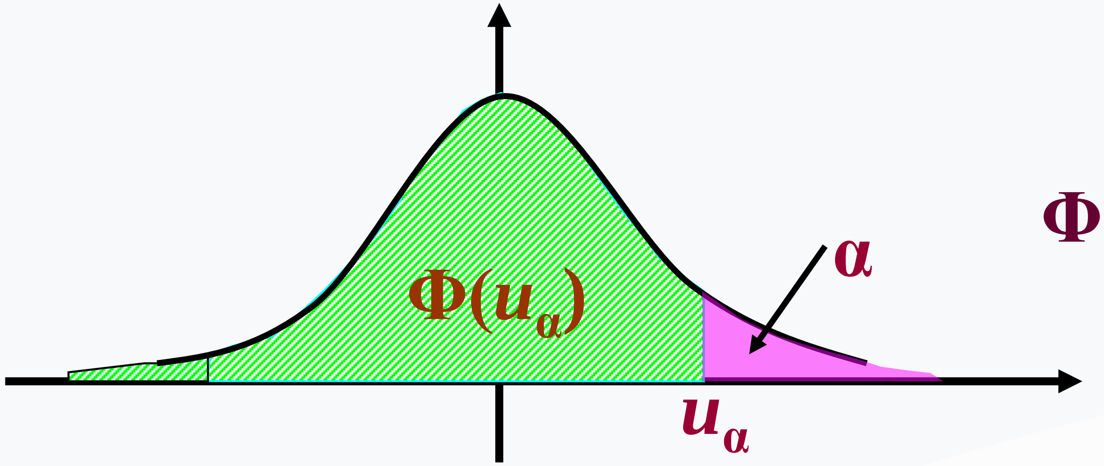
$$P\{Y>\chi_\alpha^2(n)\}=\alpha$$ (上侧分位数)

有时,我们需要求随机变量以给定概率落在某个区间上的分界点,称之为分位数.

上侧分位数 $X \sim N(0,1)$, 若实数u_α 使

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

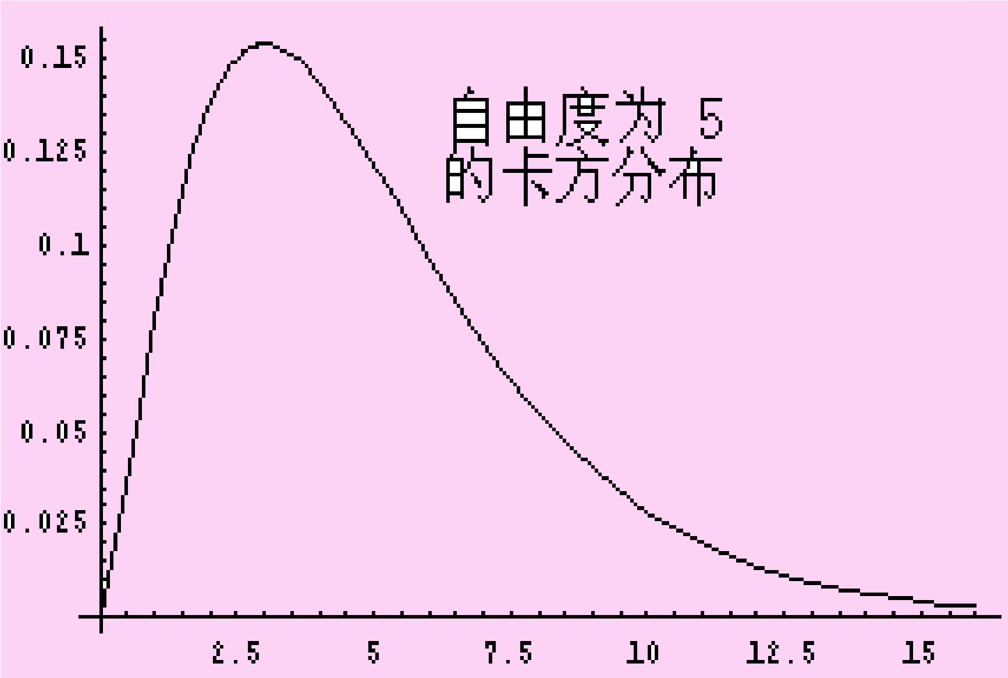
则称u_α为标准正态分布的对应于α 的上侧分位数.

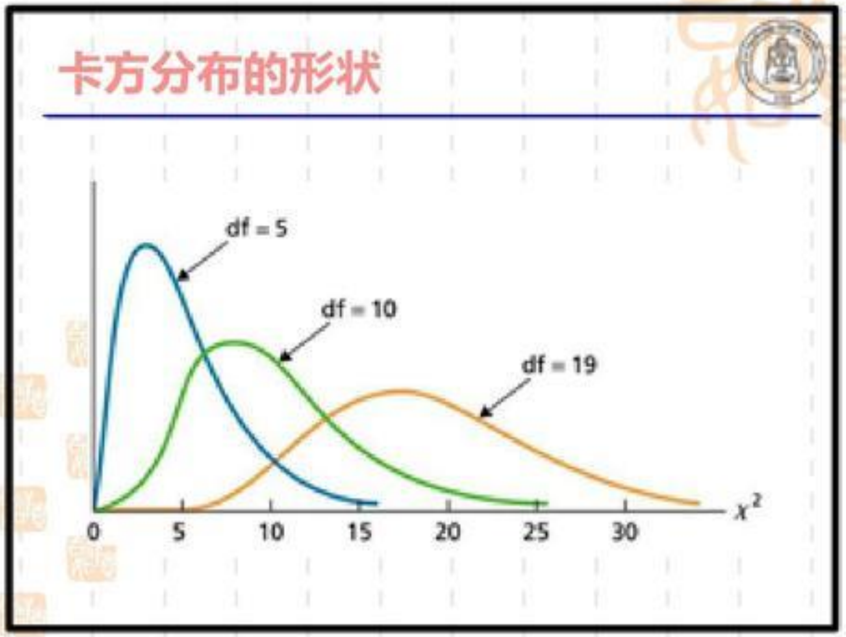


$\Phi(u_\alpha) = 1 - \alpha$

$\chi^2(n)$ 的上侧分位数$\chi^2_\alpha(n)$ ($0<\alpha<1$):

$$P\{\chi^2 > \chi^2_\alpha(n)\} = \int_{\chi^2_\alpha(n)}^{+\infty} f_{\chi^2}(x)dx = \alpha$$





也可以使用 Excel 软件点击“插入 → 函数 → 统计函数 → CHIINV(或TINV)”,然后根据对话框填入相应的数值就可获得χ^2分布(或t分布)对应的上侧分位数.

例 查表计算概率

$1.X \sim N(0,1),\quad P\{-1.58 \le X \le 1.96\} = ?$

$2.\chi^2 \sim \chi^2(15),\quad P\{6.262 \le \chi^2 \le 24.996\} = ?$

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

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

$\therefore P\{-1.58 \le X \le 1.96\}$

$= P\{X \le 1.96\} - P\{X \le -1.58\}$

$= \Phi(1.96) - [1-\Phi(1.58)]$

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

$= 0.975 - (1-0.943) = 0.918$

$$\chi^2_{0.975}(15)=6.262$$

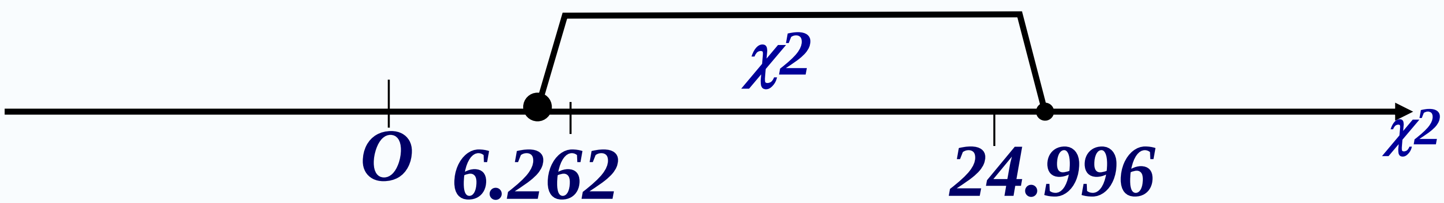
$$\chi^2_{0.05}(15)=24.996$$

2. $\chi^2 \sim \chi^2(15)$,

$$P\{6.262 \le \chi^2 \le 24.996\} = ?$$

$$= P\{\chi^2 \ge 6.262\} - P\{\chi^2 > 24.996\}$$ (上侧分位数)

$$= 0.975 - 0.05 = 0.925$$



$$P\{\chi^2 > \underline{\chi^2_\alpha(n)}\} = \int_{\chi^2_\alpha(n)}^{+\infty} f_{\chi^2}(x)dx = \underline{\alpha}$$

注意：查表时应注意分布表的定义与查法！

续例：统计量的分布(之一)

设$X_1, X_2, \ldots, X_n$是来自正态总体$X \sim N(\mu,\sigma^2)$的容量为 n 的样本, 求下列统计量的概率分布:

1. $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ 2. $Y = \dfrac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$

解： 故 $Y = \dfrac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2(n)$

当总体$X \sim N(0,0.3^2), n=10$, 求$P\{\sum_{i=1}^{10} X_i^2 > 1.44\}$

$$P\left\{\sum_{i=1}^{10} X_i^2 > 1.44\right\} = P\left\{\frac{\sum_{i=1}^{10} X_i^2}{0.3^2} > \frac{1.44}{0.3^2}\right\} = P\{\chi^2(10) > 16\}$$

$$= 0.1.$$

- χ^2分布的三条性质：

性质1.（数字特征）设$\chi^2\sim\chi^2(n)$，则有

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

证明：$\chi^2=\sum_{i=1}^{n}X_i^2$（卡方分布的结构分布）

且 $X_1, X_2, \ldots, X_n$相互独立，$X_i\sim N(0,1)$

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

$$D(\chi^2)=\sum_{i=1}^{n}D(X_i^2)=\sum_{i=1}^{n}\{E(X_i^4)-[E(X_i^2)]^2\}=2n$$

注：$X\sim N(0,\sigma^2)$，则$E\left(X^n\right)=\begin{cases}0 & n=1,3,5,\cdots\\ \sigma^n(n-1)(n-3)\cdots 1 & n=2,4,6,\cdots\end{cases}$

• χ^2分布的三条性质：

性质1.（数字特征）设$\chi^2 \sim \chi^2(n)$，则有

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

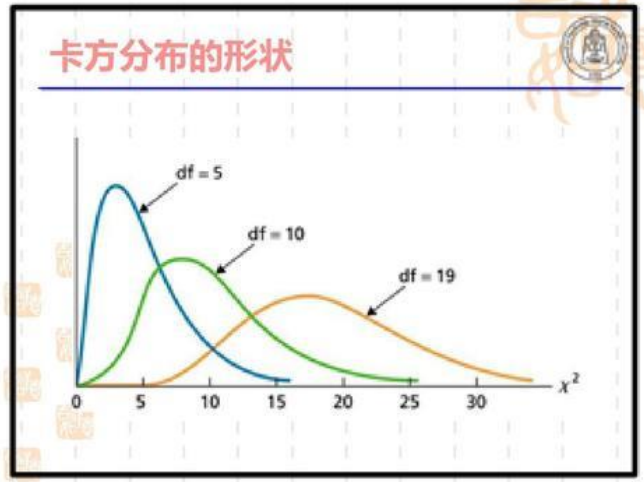
性质2.(可加性) 设Y_1、Y_2 相互独立且$Y_1 \sim \chi^2(n_1)$,

$Y_1 \sim \chi^2(n_2)$, 则 $Y_1+Y_2 \sim \chi^2(n_1+n_2)$

证明：记 $Y_1=\sum_{i=1}^{n_1} X_i^2$， $Y_2=\sum_{i=n_1+1}^{n_1+n_2} X_i^2$

且X_i，$i=1,2,\ldots,n_1+n_2$相互独立, $X_i \sim N(0,1)$,

则 $Y_1+Y_2=\sum_{i=1}^{n_1+n_2} X_i^2 \sim \chi^2(n_1+n_2)$.



• χ^2分布的三条性质:

性质1.（数字特征）设$\chi^2 \sim \chi^2(n)$, 则有

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

性质2. (可加性)设Y_1、Y_2 相互独立且$Y_1 \sim \chi^2(n_1)$, $Y_1 \sim \chi^2(n_2)$, 则 $Y_1+Y_2 \sim \chi^2(n_1+n_2)$

性质3.（大样本分位数 $\chi^2_\alpha(n)$）

当n足够大(如n>45)时, 有 $\chi^2_\alpha(n) \approx n + u_\alpha\sqrt{2n}$,

其中u_α满足$\Phi(u_\alpha) = 1-\alpha$.

$$\frac{\chi^2_\alpha(n) - n}{\sqrt{2n}} \approx u_\alpha$$

$$\chi^2 = \sum_{i=1}^{n} X_i^2 \sim N(n, 2n)$$

近似服从. (独立同分布中心极限定理)

英国威廉·戈塞特(Cosset. W.S.)于1908年首先发表，当时他还在首都柏林的健力士酿酒厂工作, 从事试验和数据分析工作(数学和化学双学位), 对误差大量感性认识.

n不大时(小样本), $T=\dfrac{\overline{X}-\mu}{S/\sqrt{n}}$ 的观察值(-1,1), (-2,2), (-3,3)频率与$N(0,1)$在这区间的概率相差较大, 怀疑不属于正态分布，专门去伦敦学习统计方法，提出t分布。

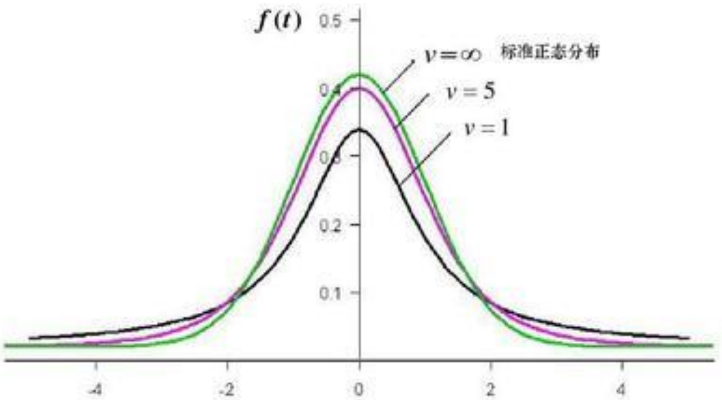


图4-2 不同自由度的 t 分布图

1.单峰分布，以0为中心，左右两侧完全对称

与老(K. Pearson), 小Pearson(英国最高水平统计学家)研究大样本统计精神违背，不被接受，只在酿酒公司使用。

因为不能以他本人的名义发表，所以论文使用了学生(Student)这一笔名. 英国Fisher(统计学的创始人:罗纳德·费雪)遇到小样本农业试验，t分布价值才被广泛接受，开创小样本统计学方法, 而正是他将此分布称为学生分布。

小故事：t 检验、啤酒、“学生”与威廉•戈瑟特

1899 年，由于爱尔兰都柏林的吉尼斯啤酒厂热衷于聘用剑桥、牛津的优秀毕业生，学化学的牛津毕业生威廉•戈瑟特 (William Gosset, 1876—1937) 到该厂就职，希望将他的生物化学知识用于啤酒生产过程. 为降低啤酒质量监控的成本，戈瑟特发明了 t 检验法，1908 年在 *Biometrika* 发表. 为防止泄漏商业机密，戈瑟特发表文章时用了笔名“学生”，于是该方法被称为“学生氏 t 检验”(Student's t-test).

吉尼斯啤酒厂是一家很有远见的企业，为保持技术人员的高水准，该厂像高校一样给予技术人员“学术假”，1906—1907 年戈瑟特得以到“统计学之父”卡尔•皮尔逊 (Karl Pearson, 1857—1936) 教授在伦敦大学学院 (University College London, 简称 UCL) 的实验室访问学习. 因此，很难说 t 检验法是戈瑟特在啤酒厂还是在 UCL 访学期间提出的，但“学生”与戈瑟特之间的联系是被 UCL 的统计学家们发现的，尤其因为皮尔逊教授恰是 *Biometrika* 的主编.

3. t 分 布(学生t-分布可简称为t分布)

(又称学生氏分布----第一个研究者以*Student*作笔名发表文章)

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

称随机变量T服从自由度为n 的t分布, 记$T\sim t(n)$.

定理6.2.2 (结构定理) 设随机变量X, Y 相互独立, $X\sim N(0,1)$, $Y\sim \chi^2(n)$, 则

$$T=\frac{X}{\sqrt{Y/n}}\sim t(n)$$

即随机变量 T 服从自由度为 n 的 t 分布.

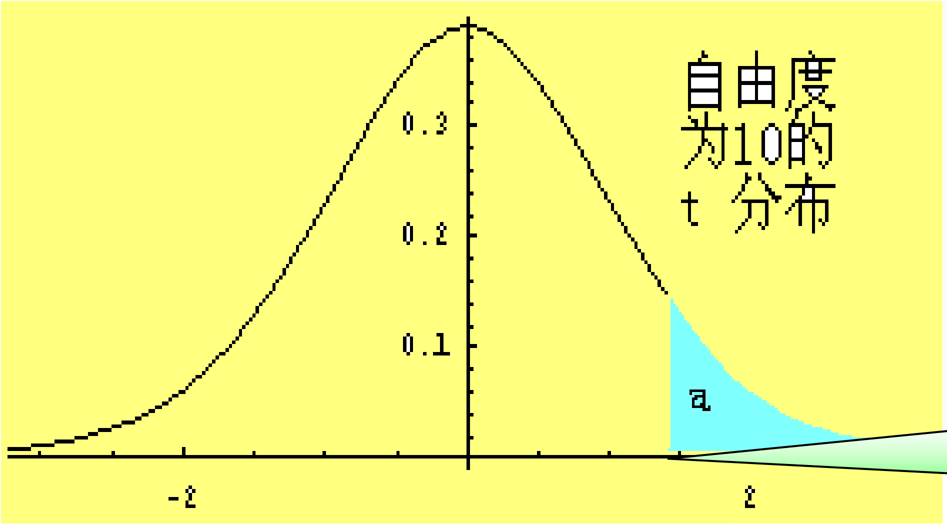
定理6.2.2(结构定理) 设随机变量X, Y相互独立, $X\sim N(0,1)$, $Y\sim\chi^2(n)$, 则

$$T=\frac{X}{\sqrt{Y/n}}\sim t(n)$$

即随机变量 T 服从自由度为 n 的 t 分布.

- $t(n)$ 的上侧分位数$t_\alpha(n)$($0<\alpha<1$):

$$P\{T>t_\alpha(n)\}=\int_{t_\alpha(n)}^{+\infty}f_T(x)dx=\alpha$$



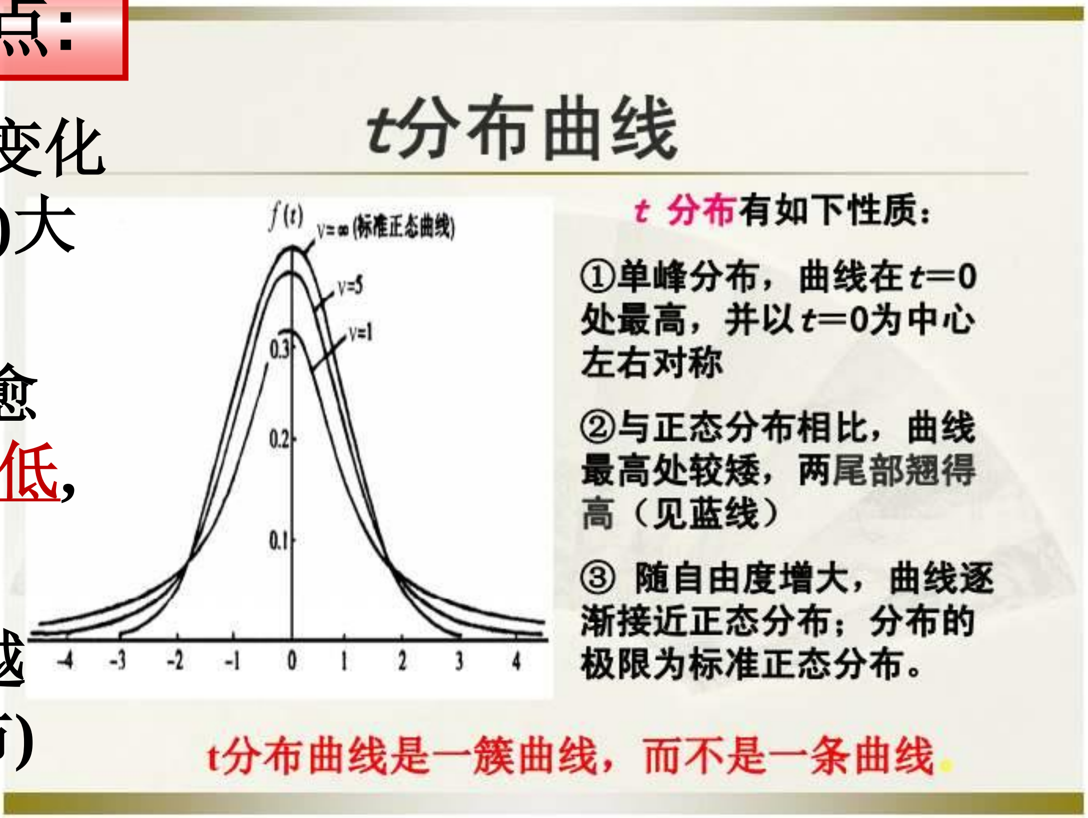
上侧分位数
$t_\alpha(n)$

T 分布密度图形的特点:

*t*分布是一簇曲线,其形态变化与*n*(确切地说与自由度*df*)大小有关.

自由度*df*越小, *t*分布曲线愈平坦(越低平),曲线中间愈低,曲线双侧尾部翘得愈高；

自由度*df*越大,*t*分布曲线越接近标准正态分布(*u*分布)曲线，如图:



1. 在置信区间估计、显著性检验等问题的计算中发挥重要作用;

2. 金融中,为一种模拟市场的好的统计分布.在外汇和股票的价格分析上很常见(重尾或后尾分布: heavy tailed).

T 分布的上侧分位数特点:

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

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

密度函数关于纵轴对称:

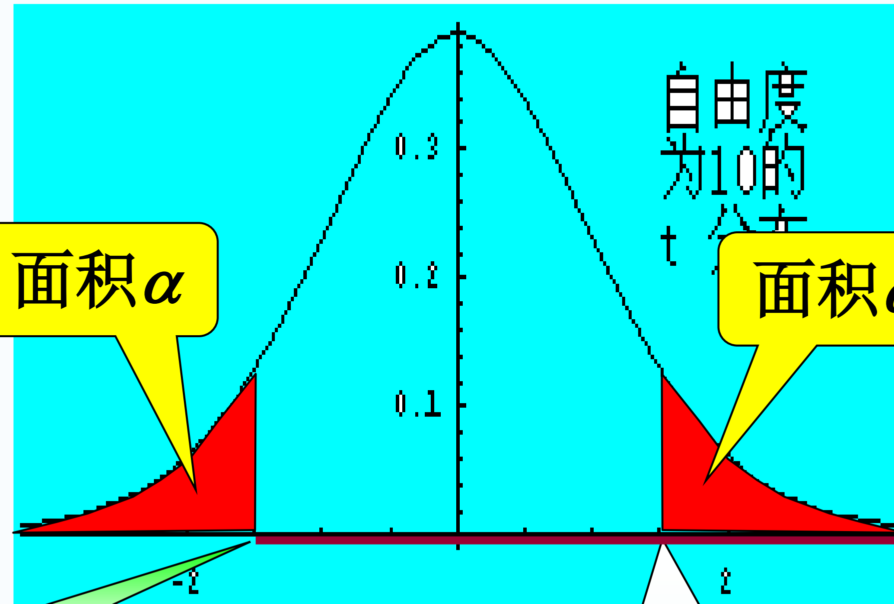
因 $\alpha=P\{T>t_\alpha\}$

$=P\{T\le -t_\alpha\}$

$=1-P\{T>-t_\alpha\}$

故 $P\{T>\underline{-t_\alpha}\}=1-\alpha$

即 $t_{1-\alpha}=-t_\alpha$



上侧分位数$t_{1-\alpha}(n)$

上侧分位数$t_\alpha(n)$

- n较大时(n>45): $t_\alpha(n)\approx u_\alpha$ $\quad T\sim N(0,1)$

T 分布上侧分位数的特点:

密度函数关于纵轴对称: $\boxed{t_{1-\alpha}(n) = -t_\alpha(n)}$

- n较大时 (n>45): $T \sim N(0,1)$

$t_\alpha(n) \approx u_\alpha$

例 查表计算: $t_{0.95}(20) = ?$ $t_{0.95}(80) = ?$

解: $\underline{t_{0.95}}(20) = t_{1-0.05}(20) = -t_{0.05}(20) = -1.7247$

$t_{0.95}(80) = -t_{0.05}\underline{(80)} \approx -u_{0.05} = -1.645$

表中$P\{T > t_\alpha(n)\} = \alpha \le 0.25,\ n \le 45$

也可以使用 Excel 软件点击"插入 → 函数 → 统计函数 → CHIINV(或TINV)",然后根据对话框填入相应的数值就可获得χ^2分布(或t分布)对应的上侧分位数.

4. F 分布

(1924年英国统计学家Ronald.A.Fisher爵士提出,并以其姓氏的第一个字母命名的)

$$f(x)=\begin{cases} n_1^{\frac{n_1}{2}} n_2^{\frac{n_2}{2}} \dfrac{\Gamma(\frac{n_1+n_2}{2})}{\Gamma(\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})} x^{\frac{n_1}{2}-1}(n_1x+n_2)^{-\frac{n_1+n_2}{2}}, & x>0 \\ 0, & x\le 0 \end{cases}$$

称F服从第一自由度为n_1, 第二自由度为n_2的F分布.

记为$F\sim F(n_1, n_2)$.

注1：一种非对称分布, 且位置不可互换.

注2：F分布有着广泛的应用，如在方差分析、回归方程的显著性检验中都有着重要的地位.

定理6.2.3 (结构定理) 设 随机变量X, Y相互独立, $X\sim\chi^2(n_1)$, $Y\sim\chi^2(n_2)$, 则

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

即随机变量 F 服从第一自由度为n_1, 第二自由度为n_2的F分布.

推论: 若 $F\sim F(n_1,n_2)$ $\Rightarrow \dfrac{1}{F}\sim F(n_2,n_1)$

$F(n_1,n_2)$的上侧分位数$F_\alpha(n_1,n_2)$ $(0<\alpha<1)$:

$$P\{F>F_\alpha(n_1,n_2)\}=\int_{F_\alpha(n_1,n_2)}^{+\infty}f_F(x)dx=\alpha$$

推论：若 $F\sim F(n_1,n_2)\Rightarrow \dfrac{1}{F}\sim F(n_2,n_1)$

推论: 若 $F\sim F(n_1,n_2)\ \Rightarrow F_\alpha(n_2,n_1)=\dfrac{1}{F_{1-\alpha}(n_1,n_2)}$

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

$\Rightarrow P\{\dfrac{1}{F}>\underline{\dfrac{1}{F_{1-\alpha}(n_1,n_2)}}\}=\alpha,$ 又因 $\dfrac{1}{F}\sim F(n_2,n_1)$

$\Rightarrow \dfrac{1}{F_{1-\alpha}(n_1,n_2)}=F_\alpha(n_2,n_1)$

$P\{\dfrac{1}{F}>F_\beta(n_2,n_1)\}=\beta$

TIPS

例 统计量的分布 (之二)

思考：$T\sim t(n)\Rightarrow T^2\sim ?;\dfrac{1}{T^2}\sim ?$

例： 统计量的分布(之二)

设 $X_1, X_2, \ldots, X_{n+m}$是来自正态总体 $X \sim N(0,\sigma^2)$的容量为 $n+m$的样本, 求下列统计量的概率分布:

$$1.\ Y=\frac{1}{\sigma^2}\sum_{i=1}^{n+m}X_i^2 \quad 2.\ Z=\sqrt{\frac{m}{n}}\sum_{i=1}^{n}X_i\Big/\sqrt{\sum_{i=n+1}^{n+m}X_i^2} \quad 3.\ \frac{1}{Z^2}$$

解:

1. $\dfrac{X_i}{\sigma} \sim N(0,1)$且所有$\dfrac{X_i}{\sigma}$相互独立$(i=1,2,\ldots,n+m)$

$$\text{故}\ Y=\frac{1}{\sigma^2}\sum_{i=1}^{n+m}X_i^2=\sum_{i=1}^{n+m}\left(\frac{X_i}{\sigma}\right)^2 \sim \chi^2(n+m)$$

正态总体 $X\sim N(0,\sigma^2)$, $X_1, \ldots, X_{n+m}$容量为 $n+m$的样本.

$$1.\ Y=\frac{1}{\sigma^2}\sum_{i=1}^{n+m}X_i^2 \qquad 2.\ Z=\sqrt{\frac{m}{n}}\sum_{i=1}^{n}X_i\Big/\sqrt{\sum_{i=n+1}^{n+m}X_i^2} \qquad 3.\ \frac{1}{Z^2}$$

2. 因$\sum_{i=1}^{n}X_i\sim N(0,n\sigma^2)$ $\Rightarrow U=\sum_{i=1}^{n}X_i\Big/\sqrt{n\sigma^2}\sim N(0,1)$

同时 $V=\sum_{i=n+1}^{n+m}\frac{X_i^2}{\sigma^2}\sim\chi^2(m)$, **U与V相互独立**

由t分布结构定理:

$$Z=\sqrt{\frac{m}{n}}\sum_{i=1}^{n}X_i\Big/\sqrt{\sum_{i=n+1}^{n+m}X_i^2}=\frac{U}{\sqrt{V/m}}\sim t(m)$$

$$1.Y=\frac{1}{\sigma^2}\sum_{i=1}^{n+m}X_i^2;\ 2.Z=\sqrt{\frac{m}{n}}\sum_{i=1}^{n}X_i\Big/\sqrt{\sum_{i=n+1}^{n+m}X_i^2};\ 3.\frac{1}{Z^2}$$

3. 因$Z\sim t(m)$, 根据 t 分布结构定理, 有

$$Z=\frac{U}{\sqrt{V/m}},\quad \text{而且}\quad U\sim N(0,1)\Rightarrow U^2\sim\chi^2(1),$$

$$V\sim\chi^2(m),\quad U\text{与}V\text{相互独立}$$

U^2与V相互独立　　根据 F分布结构定理

故 $$\frac{1}{Z^2}=\frac{V/m}{U^2/1}\sim F(m,1)$$

统计推断三个方面：1. 抽样分布(精确分布)；
2. 参数估计；(已知分布类型)
3. 假设检验。

参数估计
- 点估计
  - 矩估计
  - 极大似然估计
- 区间估计

二、正态总体的抽样分布

古代统计学的发展与天文观察数据及测地资料的处理密切相关，这时都牵涉到对随机误差的研究．高斯首先强调了正态分布在描述观察误差中的重要性，之后生物统计中也发现了正态分布是自然界中最常见的分布，加上正态分布本身的良好性质，因此数理统计学的基本理论主要是在总体服从正态分布的假定下建立起来的．

二、抽样分布定理(正态总体的统计量分布)

定理6.2.4: 设$X_1, X_2, \ldots, X_n$是正态总体$X \sim N(\mu, \sigma^2)$的样本,

$\overline{X}, S^2$分别是样本均值和样本方差, 则

Fisher 引理

(1) $\overline{X}$与S^2相互独立;

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

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

(4) $\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$ μ已知: $\sum_{i=1}^{n}(\frac{X_i-\mu}{\sigma})^2 \sim \chi^2(n)$.