

第二节 常用统计分布

- 一、常见分布
- 二、概率分布的分位数
- 三、内容小结

一、常见分布

- (1) χ^2 分布(掌握定义)

定义5.6: 设随机变量 X_1, X_2, \dots, X_n 独立同分布, 且每个 $X_i \sim N(0,1)$, 则称随机变量

$$\chi_n^2 = X_1^2 + X_2^2 + \dots + X_n^2 = \sum_{i=1}^n X_i^2$$

所服从的分布为自由度为 n 的 χ^2 分布.

记为 $\chi_n^2 \sim \chi^2(n)$. 随机变量 χ_n^2 也称为 χ^2 变量.

自由度: 独立变量的个数 n

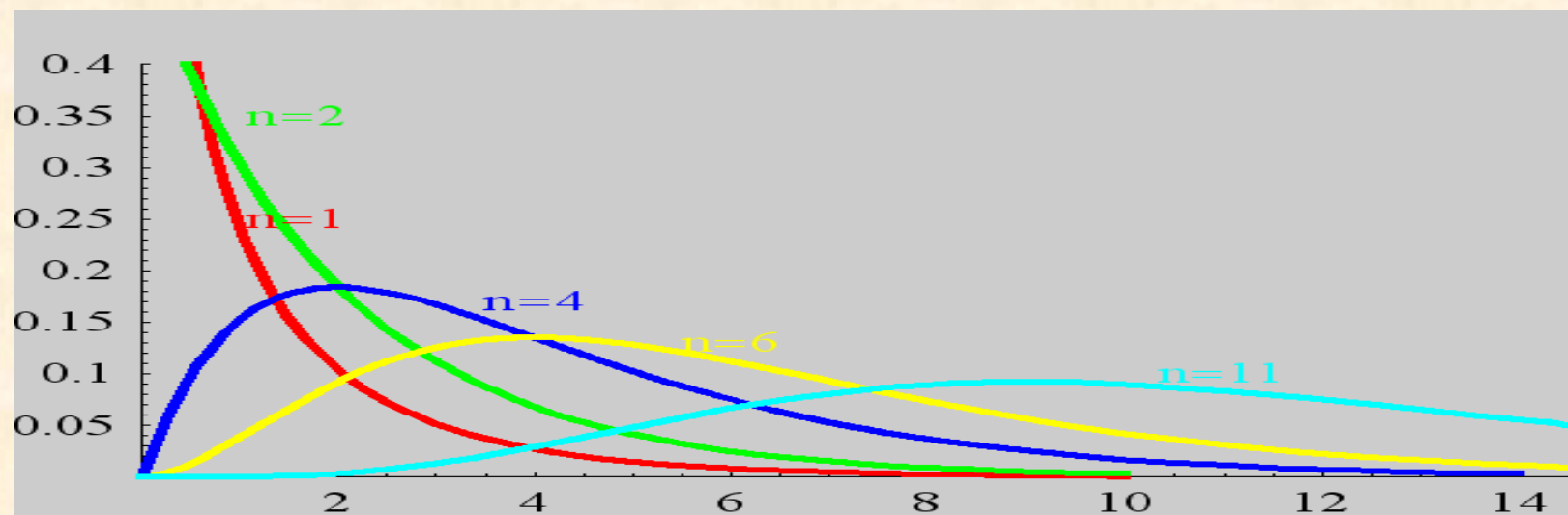
(2) χ^2 的概率密度 (了解)

定理5.4 χ^2 分布的概率密度为:

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

证: 略

$\chi^2(n)$ 分布的概率密度曲线如图.



(3) χ^2 分布的性质 (掌握)

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

性质2 (χ^2 分布的数学期望和方差)

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

证 因为 $X_i \sim N(0, 1)$, 所以

$$E(X_i^2) = D(X_i) + [E(X_i)]^2 = 1,$$

$$\begin{aligned} E(X_i^4) &= \int_{-\infty}^{+\infty} x^4 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{-2}{\sqrt{2\pi}} \int_0^{+\infty} x^3 \cdot de^{-\frac{x^2}{2}} \\ &= \frac{-2}{\sqrt{2\pi}} \left[x^3 e^{-\frac{x^2}{2}} \Big|_0^{+\infty} - \int_0^{+\infty} 3x^2 \cdot e^{-\frac{x^2}{2}} dx \right] \\ &= \frac{6}{\sqrt{2\pi}} \int_0^{+\infty} -x de^{-\frac{x^2}{2}} = \frac{-6}{\sqrt{2\pi}} \left[xe^{-\frac{x^2}{2}} \Big|_0^{+\infty} - \int_0^{+\infty} e^{-\frac{x^2}{2}} dx \right] \\ &= 3 \end{aligned}$$

$$\begin{aligned}\therefore D(X_i^2) &= E(X_i^4) - [E(X_i^2)]^2 \\ &= 3 - 1 = 2, \quad (i = 1, 2, \cdots, n)\end{aligned}$$

$$\begin{aligned}\text{故 } E(\chi_n^2) &= E\left(\sum_{i=1}^n X_i^2\right) \quad (E(X_i^2) = 1) \\ &= \sum_{i=1}^n E(X_i^2) = n,\end{aligned}$$

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

性质3 (了解)

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

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

证 由假设和定义5.6, $\chi_n^2 = \sum_{i=1}^n X_i^2$, 其中 X_1, X_2, \dots, X_n

独立且每个 $X_i \sim N(0,1)$, 因而 $X_1^2, X_2^2, \dots, X_n^2$ 独立同分布,
且

$$E(X_i^2) = 1, \quad D(X_i^2) = 2 \quad (i = 1, 2, \dots, n)$$

由中心极限定理得

$$\lim_{n \rightarrow \infty} P\left\{\frac{\chi_n^2 - n}{\sqrt{2n}} \leq x\right\} = \lim_{n \rightarrow \infty} P\left\{\frac{\sum_{i=1}^n X_i^2 - n\mu}{\sqrt{n}\sigma} \leq x\right\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

即 χ^2 分布的极限分布是正态分布,也即,当 n 很大时

$\frac{\chi_n^2 - n}{\sqrt{2n}}$ 近似服从 $N(0,1)$. 进而 $\chi_n^2 \overset{\text{近似}}{\sim} N(n, 2n)$.

例1 (重点)

设 X_1, X_2, \dots, X_6 为来自正态总体 $N(0,1)$ 的一组样本, 求 C_1, C_2 使得

$$Y = C_1(X_1 + X_2)^2 + C_2(X_3 + X_4 + X_5 + X_6)^2$$

服从 χ^2 分布.

解 $X_1 + X_2 \sim N(0,2)$, 则 $Y_1 = \frac{X_1 + X_2}{\sqrt{2}} \sim N(0,1)$

同理 $X_3 + X_4 + X_5 + X_6 \sim N(0,4)$,

则 $Y_2 = \frac{X_3 + X_4 + X_5 + X_6}{\sqrt{4}} \sim N(0,1)$

$$\text{又} \because Y_1 = \frac{X_1 + X_2}{\sqrt{2}} \text{ 与 } Y_2 = \frac{X_3 + X_4 + X_5 + X_6}{\sqrt{4}}$$

相互独立

$$\begin{aligned} \text{所以 } & \left(\frac{X_1 + X_2}{\sqrt{2}} \right)^2 + \left(\frac{X_3 + X_4 + X_5 + X_6}{\sqrt{4}} \right)^2 \\ &= Y_1^2 + Y_2^2 \sim \chi^2(2) \end{aligned}$$

则 $C_1 = 1/2, C_2 = 1/4$.

2. t 分布

(1) **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**.

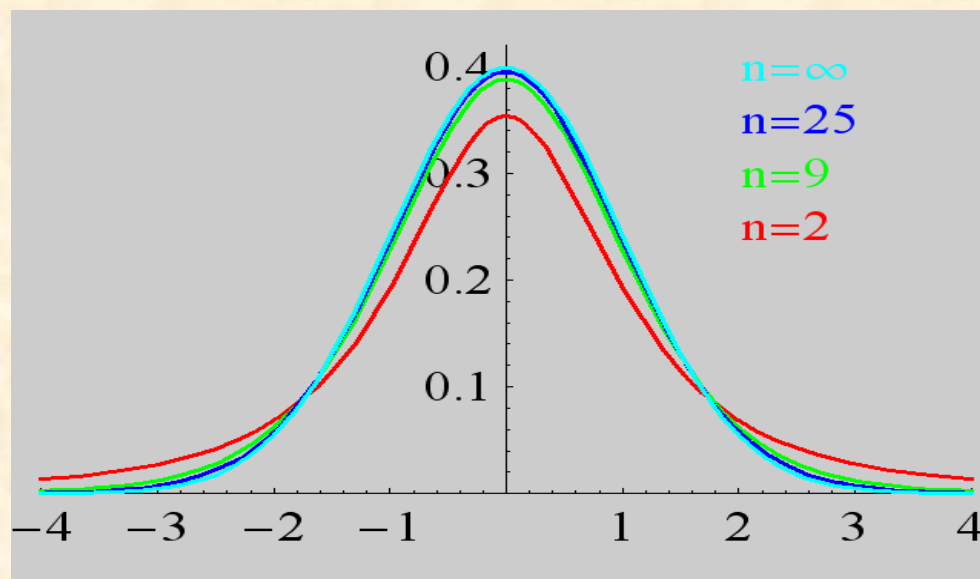
(2) $t(n)$ 分布的概率密度函数为（掌握密度函数形状）

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

t 分布的概率密度曲线如图

显然图形是关于
 $t = 0$ 对称的.

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.



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

所以当 n 足够大时 t 分布近似于 $N(0,1)$ 分布,
但对于较小的 n , t 分布与 $N(0,1)$ 分布相差很大.

(3) T 的数字特征

$$E(T) = 0,$$

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

例2 设 $X \sim N(\mu, \sigma^2)$, $\frac{Y}{\sigma^2} \sim \chi^2(n)$, 且 X, Y 相互独立,

试求 $T = \frac{X - \mu}{\sqrt{Y/n}}$ 的概率分布. (判断分布)

解 $\because X \sim N(\mu, \sigma^2), \therefore \frac{X - \mu}{\sigma} \sim N(0, 1)$

又 $\frac{Y}{\sigma^2} \sim \chi^2(n)$, 且 X, Y 独立, 则 $\frac{X - \mu}{\sigma}$ 与 $\frac{Y}{\sigma^2}$ 独立,

由定义5.7, $T = \frac{X - \mu}{\sqrt{Y/n}} = \frac{(X - \mu) / \sigma}{\sqrt{(Y / \sigma^2) / n}} \sim t(n)$

3. F 分布

(1) **定义5.8** 设 $X \sim \chi^2(n_1)$, $Y \sim \chi^2(n_2)$, 且 X, Y 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

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

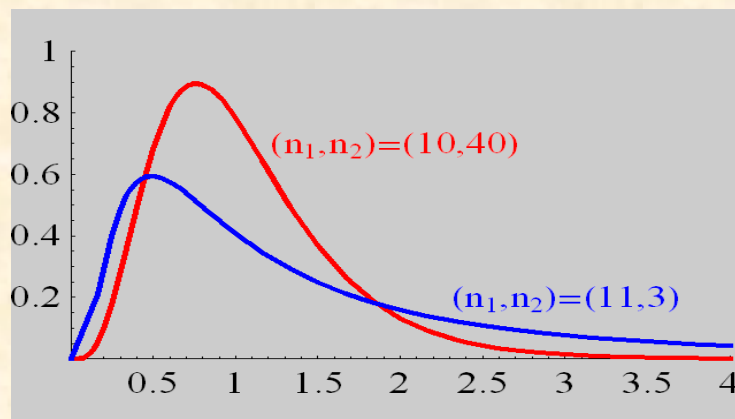
其中 n_1 称为第一自由度, n_2 称为第二自由度.

(2) $F(n_1, n_2)$ 分布的概率密度为

(了解密度
函数形状)

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

F 分布的概率密度
曲线如图



(3) F 分布有以下性质

1) 若 $F \sim F(n_1, n_2)$, 则 $\frac{1}{F} \sim F(n_2, n_1)$. 掌握

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

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

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

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

这说明 F 分布极限分布也是正态分布.

例3 已知 $T \sim t(n)$, 试证 $T^2 \sim F(1, n)$. (重点)

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 $X^2 \sim \chi^2(1)$, 且 X^2 与 Y 独立,

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

二、概率分布的分位数

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.

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	$\chi_\alpha^2(n)$	$t_\alpha(n)$	$F_\alpha(n_1, n_2)$

3. 查表法

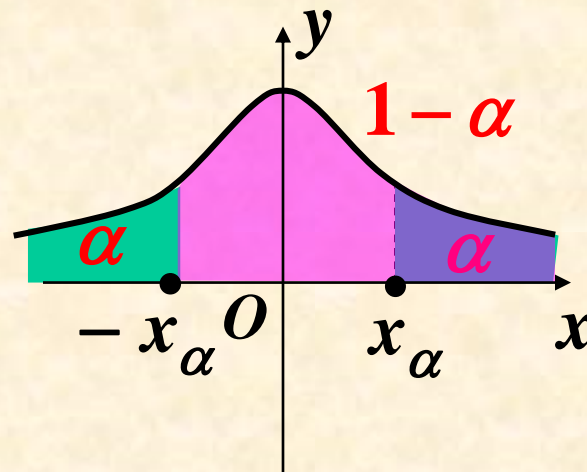
(1) 若 X 的分布密度关于 y 轴对称, 则

$$x_{1-\alpha} = -x_{\alpha}$$

特例:

1) $N(0,1): u_{1-\alpha} = -u_{\alpha}$

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



1) 正态分布的上侧分位数 u_α :

设 X 服从标准正态分布 $N(0,1)$, 则其上侧分位数 u_α 满足

$$P\{X > u_\alpha\} = \frac{1}{\sqrt{2\pi}} \int_{u_\alpha}^{+\infty} e^{-\frac{x^2}{2}} dx$$

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

即 $\Phi(u_\alpha) = 1 - \alpha$

给定 α , 由附表2可查得 u_α 的值.

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

$$u_{0.05} = 1.645,$$

附表2-1

0.95

($\alpha = 0.05$)

$$u_{0.025} = 1.96,$$

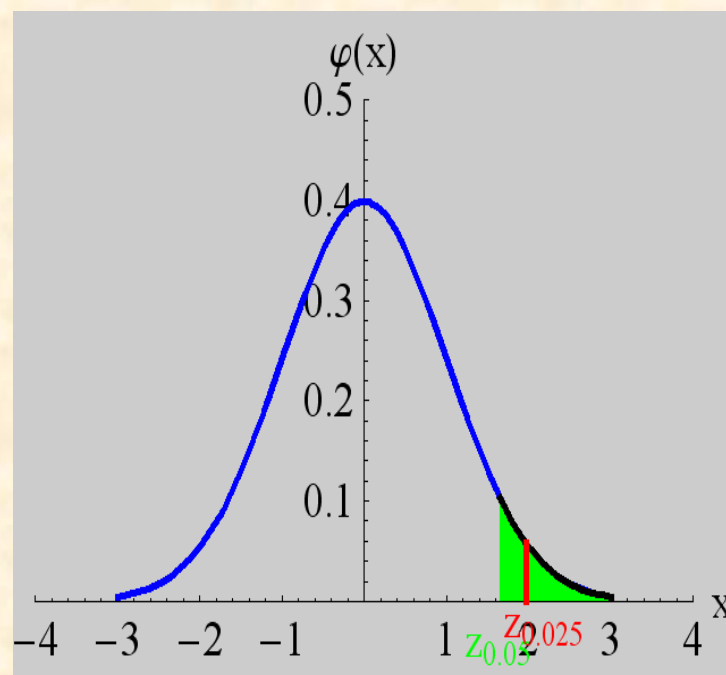
附表2-2

0.975

($\alpha = 0.025$)

根据正态分布的对称性知

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



2) t 分布的上侧分位数 $t_{\alpha}(n)$

对于给定的 α , $0 < \alpha < 1$, 称满足条件

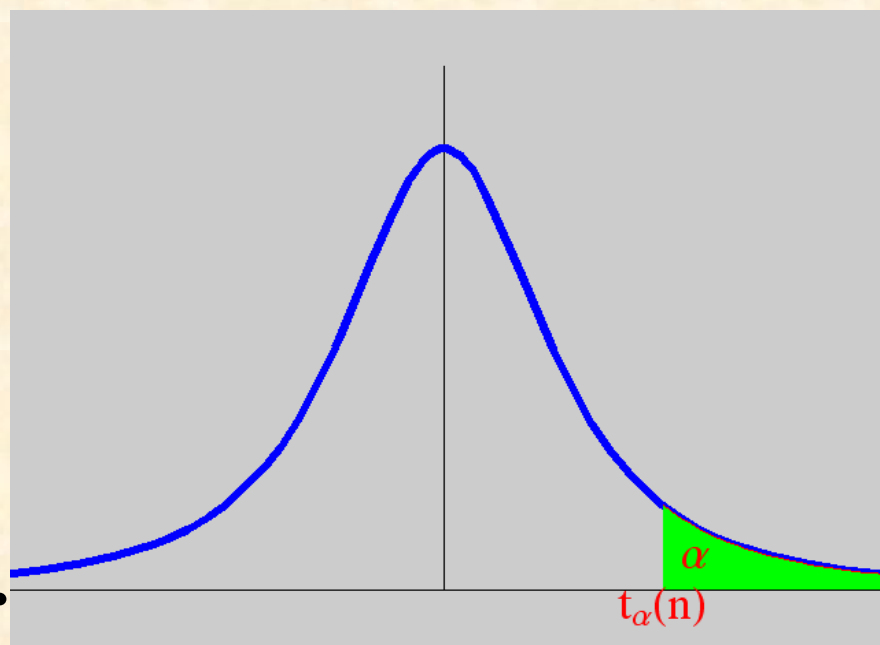
$$P\{t > t_{\alpha}(n)\} = \int_{t_{\alpha}(n)}^{\infty} h(t) dt = \alpha$$

的点 $t_{\alpha}(n)$ 为 $t(n)$ 分布的上 α 分位点.

可以通过查表求得
上 α 分位点的值.
由分布的对称性知

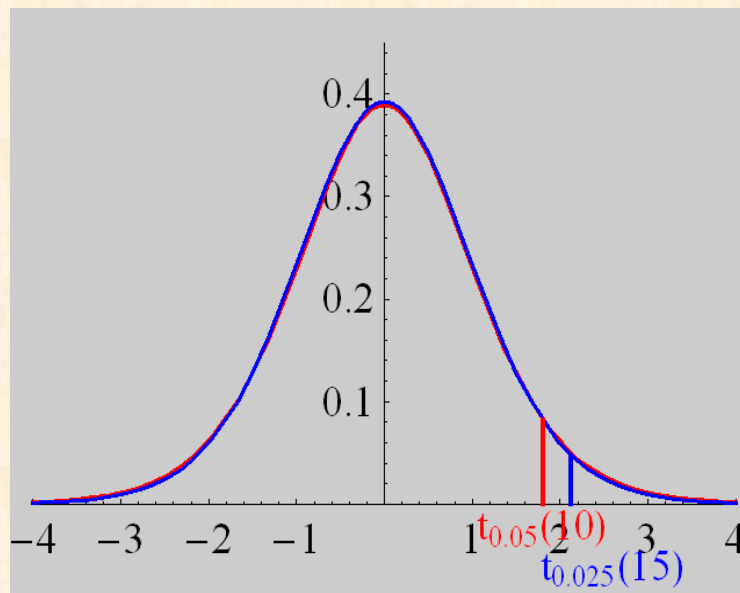
$$t_{1-\alpha}(n) = -t_{\alpha}(n).$$

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



$$t_{0.05}(10) = 1.8125, \quad \text{附表3-1}$$

$$t_{0.025}(15) = 2.1315. \quad \text{附表3-2}$$



(2) 若 X 的分布密度无对称性,

1) $\chi_{\alpha}^2(n)$: 对于给定的正数 α , $0 < \alpha < 1$, 称满足

$$P\{\chi^2 > \chi_{\alpha}^2(n)\} = \int_{\chi_{\alpha}^2(n)}^{\infty} p(y)dy = \alpha$$

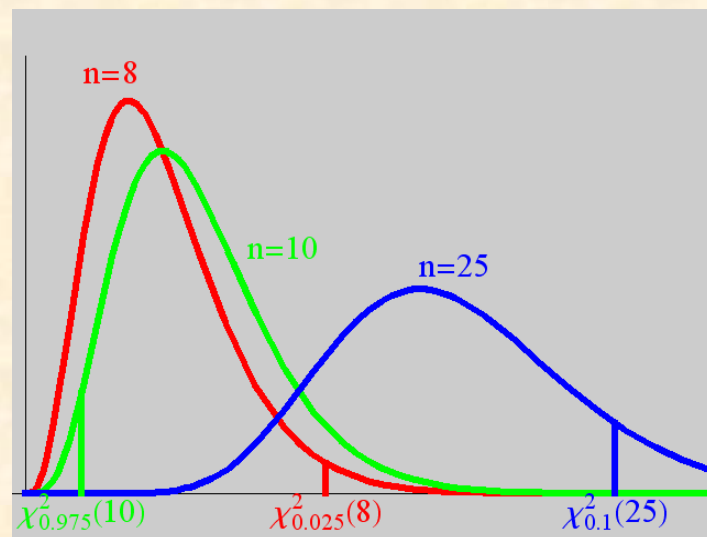
的点 $\chi_{\alpha}^2(n)$ 为 $\chi^2(n)$ 分布的上侧分位数

当 $n \leq 60$ 时, 可查表4(表4只详列到 $n=60$ 为止).

$$\chi_{0.025}^2(8) = 17.535, \quad \text{附表4-1}$$

$$\chi_{0.975}^2(10) = 3.247, \quad \text{附表4-2}$$

$$\chi_{0.1}^2(25) = 34.382. \quad \text{附表4-3}$$



当 $n > 60$ 时, $\chi^2_{\alpha}(n) \approx n + \sqrt{2n} u_{\alpha}$. (了解)

费歇(R.A.Fisher)公式:

费歇资料

当 n 充分大时, $\chi^2_{\alpha}(n) \approx n + \sqrt{2nu_{\alpha}}$.

其中 u_{α} 是标准正态分布的上 α 分位点.

例如:
$$\begin{aligned}\chi^2_{0.05}(120) &\approx 120 + \sqrt{2 \times 120} \times u_{0.05} \\ &= 120 + \sqrt{240} \times 1.64 \\ &= 145.5.\end{aligned}$$

2) $F_{\alpha}(n_1, n_2)$: 对于 $\alpha = 0.01, 0.025, 0.05, 0.1$ 等,
可直接查表5 ~ 8.

$$F_{0.05}(14, 30) = 2.31.$$

附表 5

$$F_{0.025}(7, 8) = 4.90,$$

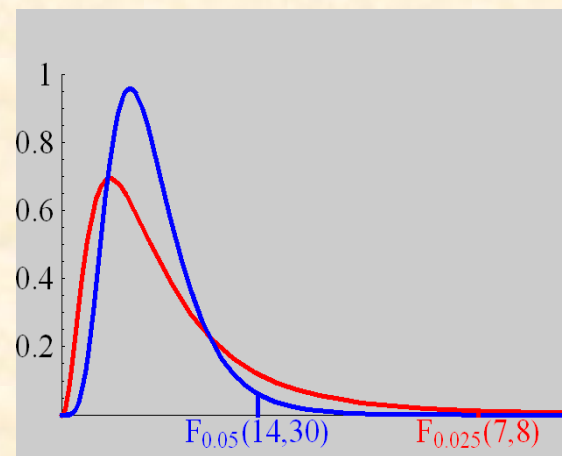
附表 8

此外, 还可利用关系

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

由 F_{α} 求得 $F_{1-\alpha}$.

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



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

证 因为 $F \sim F(n_1, n_2)$,

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

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

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

因为 $\frac{1}{F} \sim F(n_2, n_1)$,

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

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

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

辛钦定理

设随机变量 $X_1, X_2, \dots, X_n, \dots$ 相互独立, 服从同一分布, 且具有数学期望 $E(X_k) = \mu$ ($k = 1, 2, \dots$), 则对于任意正数 ε , 有 $\lim_{n \rightarrow \infty} P\left\{\left|\frac{1}{n} \sum_{k=1}^n X_k - \mu\right| < \varepsilon\right\} = 1$.

附表2-1

标准正态分布表

z	0	1	2	3	4	5	6	7	8	9
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9278	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9430	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545

附表4-1

 χ^2 分布表

n	$\alpha=0.25$	0.10	0.05	0.025	0.01	0.005
1	1.323	2.706	3.841	5.024	6.635	7.879
2	2.773	4.605	5.991	7.378	9.210	10.597
3	4.108	6.251	7.815	9.348	11.345	12.838
4	5.385	7.779	9.488	11.143	13.277	14.860
5	6.626	9.236	11.070	12.833	15.086	16.750
6	7.841	10.597	12.592	14.449	16.812	18.548
7	9.037	12.017	14.067	16.013	18.475	20.278
8	10.219	13.362	15.507	17.535	20.090	21.955
9	11.389	14.684	16.919	19.023	21.666	23.589
10	12.549	15.987	18.307	20.483	23.209	25.188
11	13.701	17.275	19.675	21.920	24.725	26.757
12	14.845	18.549	21.026	23.337	26.217	28.299
13	15.984	19.812	22.362	24.736	27.688	29.891
14	17.117	20.064	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

17.535

附表3-1

 t 分布表

n	$\alpha = 0.25$	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	2.4477	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
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.6912	1.3406	1.7531	2.1315	2.6025	2.9467
16	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208

1.8125

戈塞特



William Sealey Gosset
(1876-1937)

1899年Gosset在英国都柏林Guinness酿酒公司做酿酒师，在对小样本进行质量控制的研究中发现了t分布，即著名的Student—t分布。由于涉及商业机密，Gosset便以"学生"为笔名发表了这个研究成果。特别需要指出的是，1908年Gosset那篇论文的意义在于开创了小样本统计的新纪元。但是他数学欠佳，并没有解决t检验的理论和应用问题。他发现t分布主要依靠随机数的抽样试验(包括使用扑克牌)得出t分布曲线，因此成为用Monte-Carlo方法说明统计规律的先行者。关于t 检验理论的最后完善，Fisher, Neyman和E Pearson作出了重要贡献，或者说是理论指导实践的产物。正如后人评价的那样：“Gosset提出实际问题，Fisher和E Pearson将其转成统计问题，Neyman归纳为数学问题”。

费歇资料



Ronald Aylmer Fisher

**Born: 17 Feb 1890 in
London, England**

**Died: 29 July 1962 in
Adelaide, Australia**