第三节　　抽样分布（中）

三大抽样分布

三、抽样分布

统计三大抽样分布

统计量的分布叫做“抽样分布”。

有很多统计推断是基于正态分布的假设的，以标准正态变量为基石而构造的三个著名统计量的分布在实际中有广泛的应用，它们被称为

统计中的“三大分布”

- **χ^2分布**
- **t 分布**
- **F 分布**

1. χ^2分布

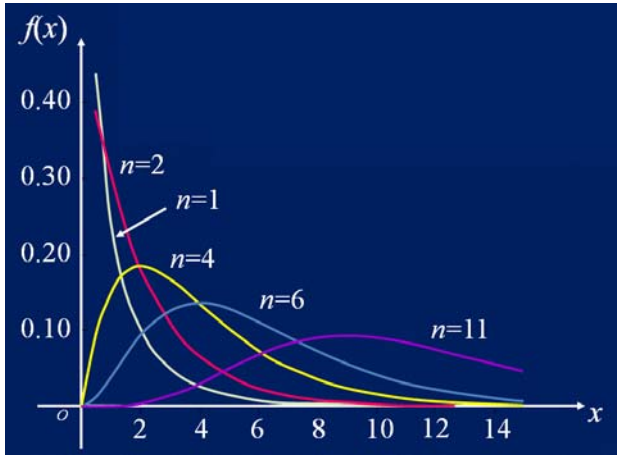
定义

设$X_1, X_2, \ldots, X_n$相互独立，都服从正态分布$N(0,1)$，则称随机变量

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

服从的分布为自由度为 n 的 χ^2 分布。

记为 $\chi^2 \sim \chi^2(n)$



χ^2分布密度函数

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

其中，伽玛函数 $\Gamma(x)$的定义是

$$\Gamma(x)=\int_0^{\infty}e^{-t}t^{x-1}dt,\quad x>0$$

【回顾】Γ分布的可加性

- 伽玛分布是统计学的一种连续概率函数。

服从参数为 α,θ 的 Γ 分布，记成 $X\sim\Gamma(\alpha,\theta)$

X的概率密度为

$$f(x)=\begin{cases}\dfrac{1}{\theta^{\alpha}\Gamma(\alpha)}x^{\alpha-1}e^{-x/\theta}, & x>0,\\ 0, & \end{cases}\qquad \alpha>0,\theta>0.$$

x<=0

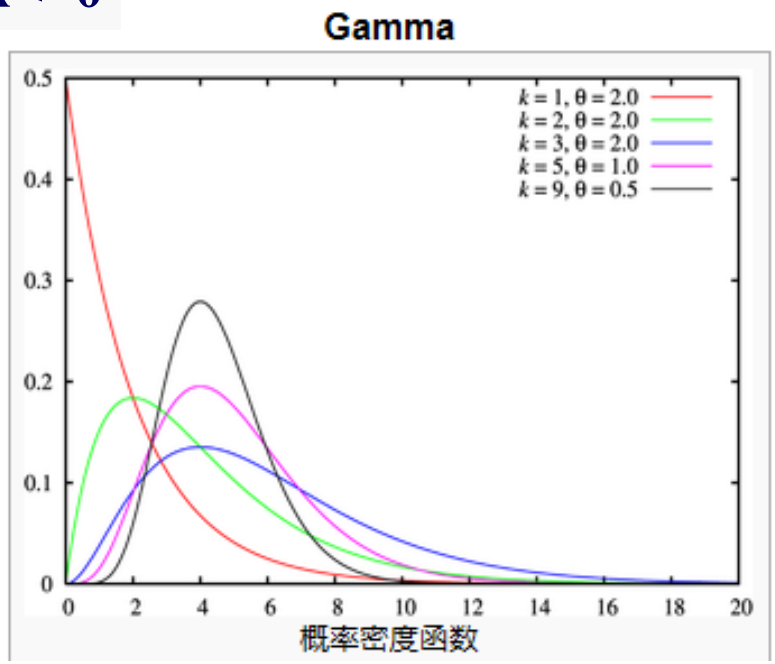
在实数域上伽玛函数定义为：

$$\Gamma(x)=\int_0^{+\infty}t^{x-1}e^{-t}\mathrm{d}t$$

- 参数α称为形状参数，θ称为尺度参数；
- 随机变量X为等到第α件事发生所需的等候时间。

其中Gamma函数之特征

$$\begin{cases}\Gamma(\alpha)=(\alpha-1)! & \text{if } \alpha \text{ is } \mathbb{Z}^+\\ \Gamma(\alpha)=(\alpha-1)\Gamma(\alpha-1) & \text{if } \alpha \text{ is } \mathbb{Z}^+\\ \Gamma\left(\frac{1}{2}\right)=\sqrt{\pi} & \end{cases}$$



χ^2分布的密度函数为

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

服从参数为 α,θ 的 Γ 分布，记成 $X\sim\Gamma(\alpha,\theta)$

X的概率密度为

$$f(x)=\begin{cases}\dfrac{1}{\theta^{\alpha}\Gamma(\alpha)}x^{\alpha-1}e^{-x/\theta}, & x>0,\\ 0, & \text{其他},\end{cases}\qquad \alpha>0,\theta>0.$$

注　已知$\chi^2(1)$就是$\Gamma\left(\frac{1}{2},2\right)$分布.由定义$X_i^2\sim\chi^2(1)$,

即$X_i^2\sim\Gamma\left(\frac{1}{2},2\right)$.再由$\Gamma$可加性知$\chi^2=\sum_{i=1}^{n}X_i^2\sim\Gamma\left(\frac{n}{2},2\right)$.

【演示】CHI分布.xls

【回顾】 Γ分布的可加性

P78 例 3 设随机变量 X,Y 相互独立，且分别服从参数为 $\alpha,\theta;\beta,\theta$ 的 Γ 分布（分别记成 $X\sim\Gamma(\alpha,\theta)$，$Y\sim\Gamma(\beta,\theta)$）.

试证明 $Z=X+Y$ 服从参数为 $\alpha+\beta,\theta$ 的 Γ 分布，即 $X+Y\sim\Gamma(\alpha+\beta,\theta)$.

χ^2分布的性质

1. 设 $X_1 \sim \chi^2(n_1)$, $X_2 \sim \chi^2(n_2)$ 且 X_1, X_2 相互独立，

   则 $X_1 + X_2 \sim \chi^2(n_1+n_2)$ χ^2 分布的可加性

   推广：设 $X_1, X_2, \ldots, X_n$ 相互独立，且 $X_i \sim \chi^2(n_i)$，则

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

2. 设 $X_1, X_2, \ldots, X_n$ 相互独立，都服从正态分布 $N \sim (\mu, \sigma^2)$，则

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

χ^2分布的性质

3. 若 $\chi^2 \sim \chi^2(n)$，χ^2 分布的数学期望与方差存在，

$E(\chi^2)=n$，$D(\chi^2)=2n$.

由$X_i \sim N(0,1)$ 有$E(X_i^2)=D(X_i)=1$

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

又$D(X_i^2)=E(X_i^4)-[E(X_i^2)]^2=3-1=2$

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

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

分部积分

$$=-\frac{1}{\sqrt{2\pi}}\underbrace{x^3\cdot e^{-x^2/2}\Big|_{-\infty}^{\infty}}_{=0}+\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}3x^2e^{-x^2/2}dx$$

$$=\frac{3}{\sqrt{2\pi}}\int_{-\infty}^{\infty}x^2e^{-x^2/2}dx=3\cdot D(X_i)=3.$$

【结论】已知X~N(0,1)，

若k=2t+1（奇数），则$E(X^k)=0$；

若k=2t（偶数），则$E(X^k)=(2t-1)(2t-3)\ldots3*1$。

定理1（独立同分布下的中心极限定理）

设随机变量$X_1, X_2, \cdots X_n$相互独立，服从同一分布，

且具有数学期望和方差：$E(X_k)=\mu, D(X_k)=\sigma^2, k=1,2,\cdots,n$，

则 $\sum_{k=1}^{n} X_k \overset{\text{近似地}}{\sim} N(n\mu, n\sigma^2)$ 即 $\dfrac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \overset{\text{近似地}}{\sim} N(0,1).$

由卡方分布的定义可知，$X=\chi^2=\sum_{i=1}^{n} X_i^2$ ，其中$X_i \sim N(0,1)$

且$X_1, X_2, \cdots, X_n$相互独立 $\Rightarrow X_1^2, X_2^2, \cdots, X_n^2$相互独立

又$E(X_i^2)=1$ ，$D(X_i^2)=2$ ，

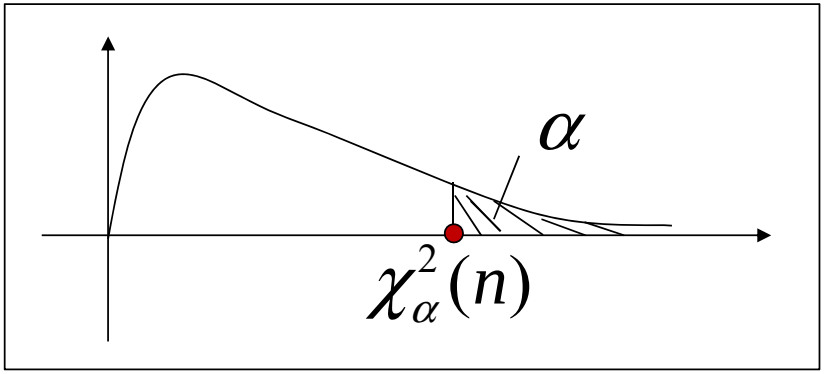
由中心极限定理可得： $\dfrac{\sum_{k=1}^{n} X_i^2 - n}{\sqrt{2n}} \overset{\text{近似地}}{\sim} N(0,1)$

χ^2分布的性质

5. 对于给定的正数$\alpha, 0<\alpha<1$,称满足条件

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

的点$\chi_\alpha^2(n)$为$\chi^2(n)$分布的上α分位点。



$\chi_\alpha^2(n)$可通过查表χ^2分布表来求得。

- P386 附表5只详细到n=40为止.

费希尔（R. A. Fisher）曾证明：

当n充分大时，近似地有

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

其中，Z_{α}是标准正态分布的上α 分位点。

利用上式可求得当 n >40 时，$\chi^{2}(n)$ 分布的上 α 分位点的近似值。

如： $\chi_{0.05}^{2}(50) \approx \frac{1}{2}(z_{0.05} + \sqrt{99})^{2} = \frac{1}{2}(1.645 + \sqrt{99})^{2} = 67.221$

2. $\boldsymbol{t}$ 分布

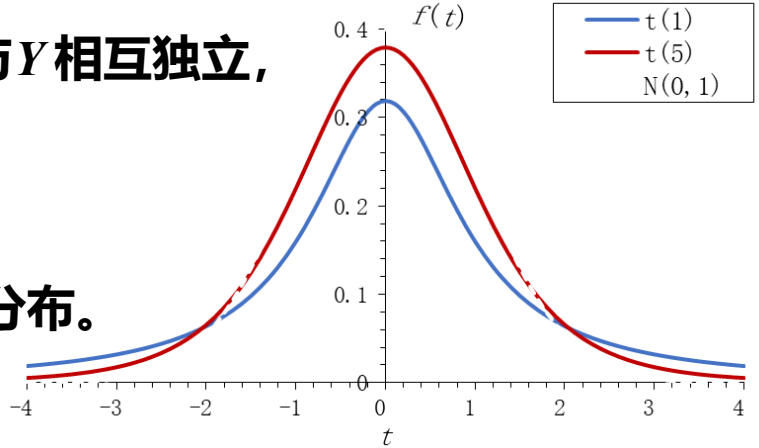
定义

设 $X \sim N(0,1)$，$Y \sim \chi^2(n)$，且 X 与 Y 相互独立，

则称变量 $$t = \frac{X}{\sqrt{Y/n}}$$

所服从的分布为自由度为 n 的 t 分布。

记为 $t \sim t(n)$.



t分布又称为学生氏分布，由英国统计学家 *William Gosset* (1876～1937) 于1908年以笔名 *student* 首次发表。

$t(n)$分布的概率密度函数为：

【演示】t分布.xls

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

特点：关于y轴对称（偶函数）；随着自由度额逐渐增大，密度曲线逐渐接近于标准正态密度曲线。

设X_1,X_2,X_3,X_4是来自正态总体$N(0,2^2)$的简单随机样本，则

$$\frac{\sqrt{3}X_1}{\sqrt{X_2^2+X_3^2+X_4^2}}$$ 服从__$t(3)$__分布；

$$\because \frac{X_i}{2} \sim N(0,1), i=1,2,3,4,$$

$$\therefore (\frac{X_2}{2})^2+(\frac{X_3}{2})^2+(\frac{X_4}{2})^2 \sim \chi^2(3)$$

故 $$\frac{\frac{X_1}{2}}{\sqrt{\{(\frac{X_2}{2})^2+(\frac{X_3}{2})^2+(\frac{X_4}{2})^2\}/3}} \sim t(3)$$

t 分布的性质

1. t分布的密度函数关于$t=0$对称。

   当n充分大时，由Γ函数的性质有

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

   即当n足够大时，$t \overset{\text{近似}}{\sim} N(0,1)$.

   当n=1时，$h(t) = \dfrac{1}{\pi(1+x^2)}$ ，即柯西分布。

2. 具有自由度为n的t分布$t \sim t(n)$ 其数学期望与方差为：

   $E(t)=0 \ (n>1)$, $D(t)=n/(n-2) \quad (n>2)$

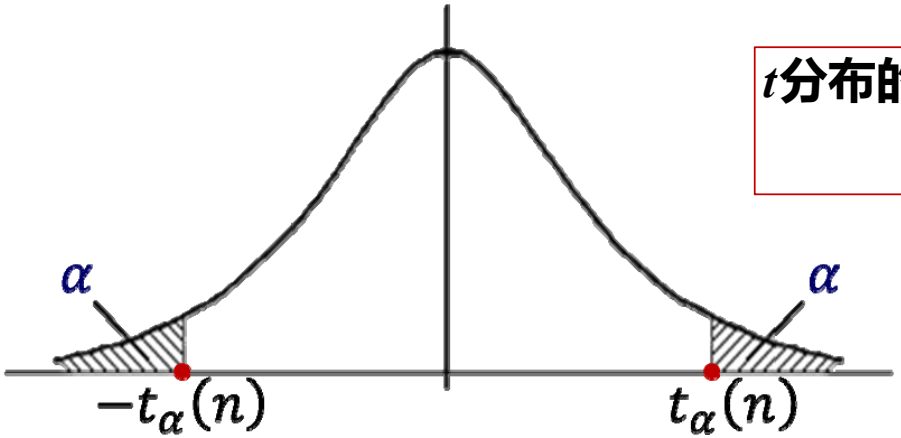
t 分布的性质

3. t分布的分位点

对于给定的$\alpha(0<\alpha<1)$ 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。



t分布的上α分位点的性质：

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

$t_\alpha(n)$可查t分布表，

例如 $t_{0.025}(15)$

$=2.1315$

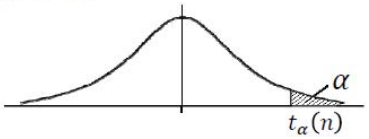
又如 $t_{0.9}(10)$

$= -t_{0.1}(10)$

$= -1.3722$

附表 4　t 分布表

$P\{t(n) > t_\alpha(n)\} = \alpha$



n \ α	0.2	0.15	0.1	0.05	0.025	0.01	0.005
1	1.376	1.963	3.0777	6.3138	12.7062	31.8207	63.6574
2	1.061	1.386	1.8856	2.9200	4.3027	6.9646	9.9248
3	0.978	1.250	1.6377	2.3534	3.1824	4.5407	5.8409
4	0.941	1.190	1.5332	2.1318	2.7764	3.7469	4.6041
5	0.920	1.156	1.4759	2.0150	2.5706	3.3649	4.0322
6	0.906	1.134	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.896	1.119	1.4149	1.8946	2.3646	2.9980	3.4995
8	0.889	1.108	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.883	1.100	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.879	1.093	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.876	1.088	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.873	1.083	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.870	1.079	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.868	1.076	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.866	1.074	1.3406	1.7531	2.1315	2.6025	2.9467
16	0.865	1.071	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.863	1.069	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.862	1.067	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.861	1.066	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.860	1.064	1.3253	1.7247	2.0860	2.5280	2.8453

当$n > 45$时，可用标准正态分布来近似，

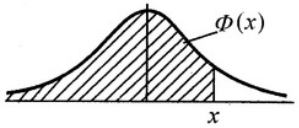
即 $t_\alpha(n) \approx z_\alpha$

例如 $t_{0.25}(50) \approx z_{0.25} \approx 0.675$

即$\Phi(?) = 0.75 \Rightarrow x \approx 0.675$

附表 2 标准正态分布表

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$



x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
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	0.7704	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

3. $\boldsymbol{F}$分布

定义

设 $U \sim \chi^2(n_1), V \sim \chi^2(n_2)$，$U$ 与 V 相互独立，

则称随机变量 $F = \frac{U/n_1}{V/n_2}$ 服从自由度为 n_1 及 n_2 的 F 分布，

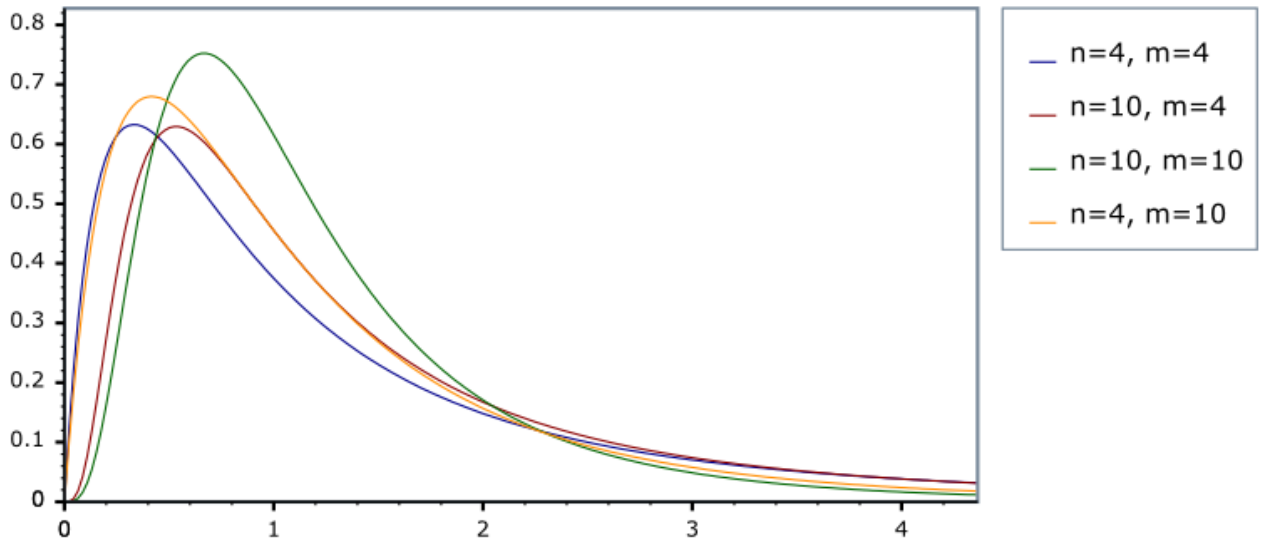
n_1 称为第一自由度，n_2 称为第二自由度，记作 $F \sim F(n_1, n_2)$ 。

F 分布由英国统计学家"R. A. Fisher"（1890~1962）于1924年提出。

由定义可见，$\frac{1}{F} = \frac{V/n_2}{U/n_1} \sim F(n_2, n_1)$

若 $F \sim F(n_1, n_2)$， F 的概率密度为

$$\psi(y) = \begin{cases} \dfrac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)} \left(\dfrac{n_1}{n_2}\right)^{\frac{n_1}{2}} (y)^{\frac{n_1}{2}-1} \left(1+\dfrac{n_1}{n_2}y\right)^{-\frac{n_1+n_2}{2}} & y > 0 \\ 0 & y \le 0 \end{cases}$$



【演示】F分布.xls

例题

例 设X_1, X_2, X_3, X_4是来自正态总体 $N(0, 2^2)$ 的简单随机样本，下列统计量服从什么分布?

(1) $\boxed{\dfrac{\sqrt{3}X_1}{\sqrt{X_2^2+X_3^2+X_4^2}}}$ ； (2) $(X_1^2+X_2^2)/(X_3^2+X_4^2)$

解： (1) $\because \dfrac{X_i}{2} \sim N(0,1), i=1,2,3,4,\quad \therefore (\dfrac{X_2}{2})^2+(\dfrac{X_3}{2})^2+(\dfrac{X_4}{2})^2 \sim \chi^2(3)$

故 $\boxed{\dfrac{\dfrac{X_1}{2}}{\sqrt{\{(\dfrac{X_2}{2})^2+(\dfrac{X_3}{2})^2+(\dfrac{X_4}{2})^2\}/3}}} \sim t(3)$

例题

例 设X_1, X_2, X_3, X_4是来自正态总体 $N(0, 2^2)$ 的简单随机样本，下列统计量服从什么分布？

(1) $\dfrac{\sqrt{3}X_1}{\sqrt{X_2^2+X_3^2+X_4^2}}$ ； (2) $\boxed{(X_1^2+X_2^2)/(X_3^2+X_4^2)}$

解： (2) $\because (\dfrac{X_1}{2})^2+(\dfrac{X_2}{2})^2 \sim \chi^2(2),\quad (\dfrac{X_3}{2})^2+(\dfrac{X_4}{2})^2 \sim \chi^2(2)$

$$\therefore \boxed{\frac{\{(\frac{X_1}{2})^2+(\frac{X_2}{2})^2\}/2}{\{(\frac{X_3}{2})^2+(\frac{X_4}{2})^2\}/2}} \sim F(2,2)$$

例 设随机变量$\boldsymbol{X \sim t(n)(n>1), Y=\frac{1}{X^2}}$，问$\boldsymbol{Y}$服从什么分布？

解： 由t分布的定义可知，$X=\frac{A}{\sqrt{B/n}} \sim t(n)$，

其中$A \sim N(0,1)$，$B \sim \chi^2(n)$且A与B相互独立

又$A^2 \sim \chi^2(1)$，故$Y=\frac{1}{X^2}=\frac{B/n}{A^2} \sim F(n,1)$

F 分布的性质

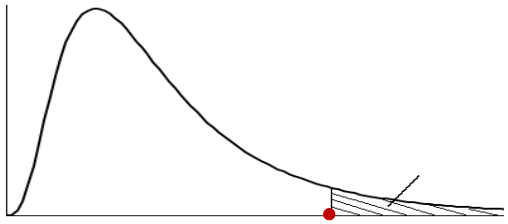
1. $\frac{1}{F}=\frac{V/n_2}{U/n_1}\sim F(n_2,n_1)$

2. F 分布的数学期望为：

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

即它的数学期望并不依赖于第一自由度n_1。

F 分布的性质



2. F 分布的分位点

对于给定的α, $0<\alpha<1$,称满足条件

$$p\{F>F_\alpha(n_1,n_2)\}=\int_{F_\alpha(n_1,n_2)}^{\infty}\psi(y)dy=\alpha$$

的点$F_\alpha(n_1,n_2)$为$F\ (n_1,n_2)$分布的上α分位点。

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

F分布的上α分位点可查表求得。

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

F 分布的性质

2. F 分布的分位

对于给定的α

$p\{F > F_\alpha($

的点$F_\alpha(n_1, n_2$

F分布的上α

$(\alpha = 0.05)$

$n_2 \backslash n_1$	1	2	3	4	5	6	7	8	9	10	12
1	161	200	216	225	230	234	237	239	241	242	244
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	1.90	2.85	2.79
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48

F分布的上α分位点可查表求得。

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

三大抽样分布

	定　义	性质	期望	方差	上α分位性质
卡方分布	若$X_1,\cdots,X_n \sim N(0,1)$，且相互独立，则$\chi^2=\sum_{i=1}^{n}X_i^2 \sim \chi^2(n)$	可加性 若$X \sim \chi^2(n)$，则当n充分大时，$\frac{X-n}{\sqrt{2n}}$近似$N(0,1)$	n	$2n$	当n>40时，$\chi_\alpha^2(n) \approx \frac{1}{2}(z_\alpha+\sqrt{2n-1})^2$
T分布	若$X \sim N(0,1)$，$Y \sim \chi^2(n)$，且X与Y相互独立 则$t=\frac{X}{\sqrt{Y/n}} \sim t(n)$	关于t=0对称 当n足够大时，$t \overset{近似}{\sim} N(0,1)$.	0	$\frac{n}{n-2}$	$t_{1-\alpha}(n)=-t_\alpha(n)$ 当$n>45$时，$t_\alpha(n) \approx z_\alpha$
F分布	若$X \sim \chi^2(n_1)$，$Y \sim \chi^2(n_2)$，且X与Y相互独立 则$F=\frac{U/n_1}{V/n_2} \sim F(n_1,n_2)$	$\frac{1}{F}=\frac{V/n_2}{U/n_1} \sim F(n_2,n_1)$	$\frac{n_2}{n_2-2}$	-----	$F_{1-\alpha}(n_1,n_2)=\frac{1}{F_\alpha(n_2,n_1)}$

罗纳德·费雪Sir Ronald Aylmer Fisher（1890~1962）

- 英国数学家和生物学家，现代统计学与现代演化论的奠基者之一
  - 安德斯·哈尔德称他是“一位几乎独自建立现代统计科学的天才”；
  - 理查·道金斯则认为他是“达尔文最伟大的继承者”。

【学术成就】

① 用亲属间的相关说明了连续变异的性状可以用孟德尔定律来解释，从而解决了遗传学中孟德尔学派和生物统计学派的论争。

② 论证了**方差分析的原理和方法**，并应用于试验设计，阐明了最大似然性方法以及随机化、重复性和统计控制的理论，指出自由度的重要性，还阐明了各种相关系数的抽样分布，亦进行过**显著性测验研究**。

③ 他提出的一些数学原理和方法对**人类遗传学、进化论和数量遗传学**的基本概念以及农业、医学方面的试验均有很大影响。