哈爾濱工業大學

第六章 数理统计的基本概念

χ^2分布、t分布和F分布

数理统计的三大分布

一、χ^2分布

χ^2 分布是由正态分布派生出来的一种分布.

定义 设 $X_1, X_2, \cdots, X_n$ 相互独立, 都服从标准正态分布 $N(0,1)$, 则称随机变量

$$Y = \sum_{i=1}^{n} X_i^2$$

服从自由度为n的 χ^2分布，记为 $Y \sim \chi^2(n)$.

自由度是指Y右端包含独立变量的个数.

χ^2分布的概率密度函数

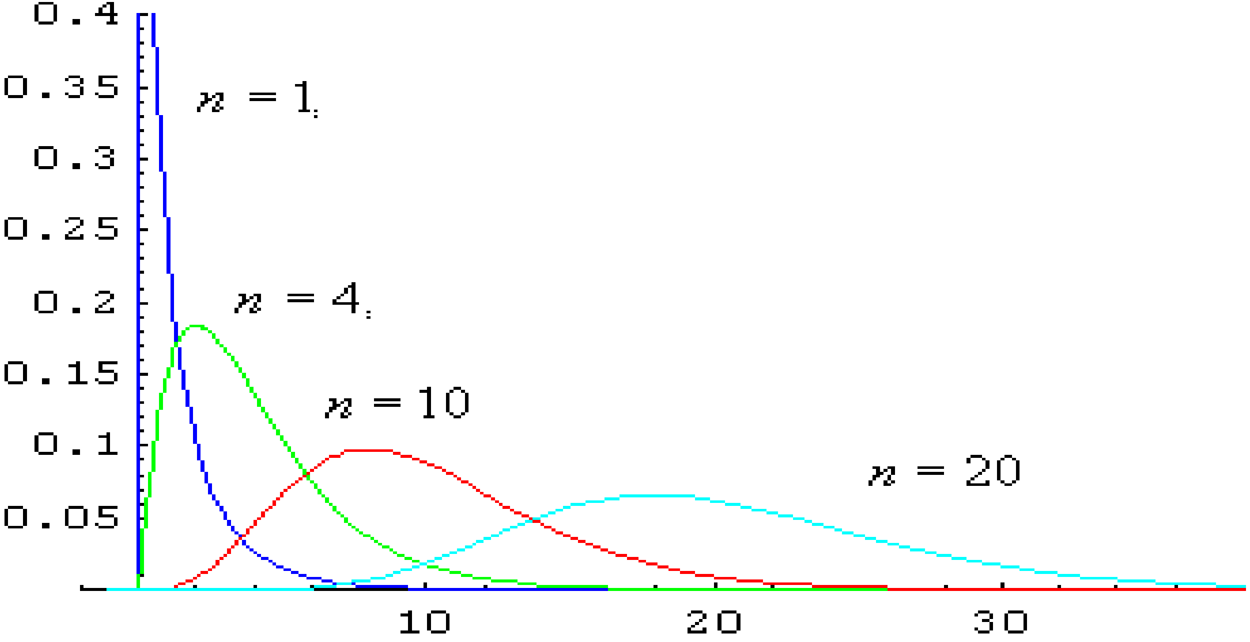
$$f(x;n)=\begin{cases}\dfrac{1}{2^{n/2}\Gamma(n/2)}x^{\frac{n}{2}-1}e^{-\frac{x}{2}}, & x\geq 0,\\ 0, & x<0.\end{cases}$$

其中

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

由阿贝(Abbe) 于1863年首先给出，后来由海尔墨特(Hermert)和皮尔逊(Pearson)分别于1875年和1900年推导出来.

χ^2分布的概率密度曲线



χ^2分布的性质

1. 设 $X \sim \chi^2(n)$，则$E(X)=n,\ D(X)=2n.$

证 设$X_1, X_2, \cdots, X_n$独立，都服从$N(0,1)$. 则

$$Y=\sum_{i=1}^{n} X_i^2 \sim \chi^2(n),\ \text{故}X\text{与}Y\text{同分布，可得}$$

$$E(X)=E(Y)=\sum_{i=1}^{n} E(X_i^2)=\sum_{i=1}^{n} D(X_i)=n,$$

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

$$=n\left[\int_{-\infty}^{+\infty} x^4 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx-1\right]=2n.$$

χ^2分布的可加性

2. 设$X \sim \chi^2(m), Y \sim \chi^2(n)$,且$X$与$Y$独立，则$X+Y \sim \chi^2(m+n)$.

推广

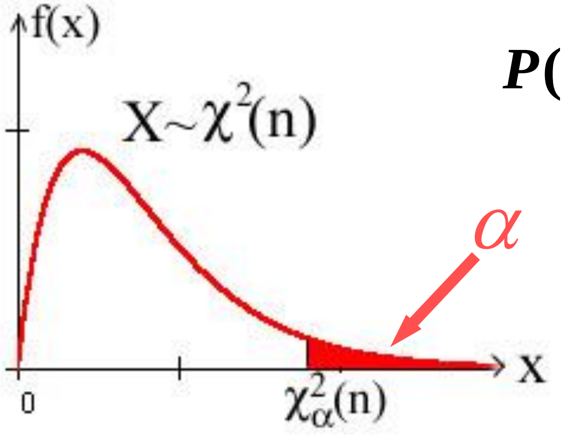
设$X_1, X_2, \cdots, X_n$独立，且$X_i \sim \chi^2(n_i)$,则$\sum_{i=1}^{n} X_i \sim \chi^2(\sum_{i=1}^{n} n_i)$.

χ^2分布的上侧α分位数或临界值

设$\alpha(0<\alpha<1)$,称满足等式 $P(X \geq \chi_\alpha^2(n)) = \alpha$

的点 $\chi_\alpha^2(n)$为$\chi^2(n)$分布的上侧α分位数或 临界值.

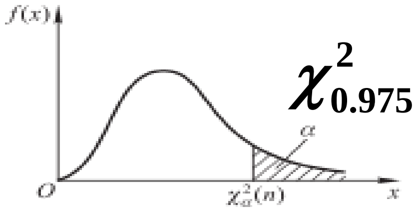
$\chi_\alpha^2(n)$的值可查χ^2分布表获得.



$$P(X \geq \chi_\alpha^2(n)) = \int_{\chi_\alpha^2(n)}^{+\infty} f(x,n)dx = \alpha$$

$$\chi_{0.025}^2(3) = 9.348,$$

附表 3 χ^2 分布表



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

$$\chi^2_{0.975}(3) = 0.216.$$

n	α=0.995	0.99	0.975	0.95	0.90	0.75
1	—	—	0.001	0.004	0.016	0.102
2	0.010	0.020	0.051	0.103	0.211	0.575
3	0.072	0.115	0.216	0.352	0.584	1.213
4	0.207	0.297	0.484	0.711	1.064	1.923
5	0.412	0.554	0.831	1.145	1.610	2.675
6	0.676	0.872	1.237	1.635	2.204	3.455
7	0.989	1.239	1.690	2.167	2.833	4.255
8	1.344	1.646	2.180	2.733	3.490	5.071
9	1.735	2.088	2.700	3.325	4.168	5.899
10	2.156	2.558	3.247	3.940	4.865	6.737
11	2.603	3.053	3.816	4.575	5.578	7.584
12	3.074	3.571	4.404	5.226	6.304	8.438

例1设$X_1,X_2,\cdots,X_n$相互独立, 都服从正态分布$N(\mu,\sigma^2)$,则

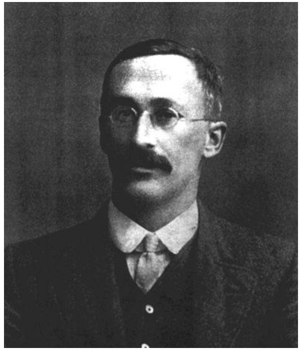
$$\chi^2=\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2\sim\chi^2(n).$$

证 由 $X_i\sim N(\mu,\sigma^2)(i=1,\cdots,n)$,则

$$Y_i=\frac{X_i-\mu}{\sigma}\sim N(0,1),\text{且}Y_1,Y_2,\cdots Y_n\text{独立，从而}$$

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

t 分布



William Gosset
1908提出了t分布.

■ 定义 设X,Y相互独立，且

$X\sim N(0,1),\ Y\sim\chi^2(n)$,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$T\sim t(n)$.

也称为学生氏分布.

t 分布的概率密度函数

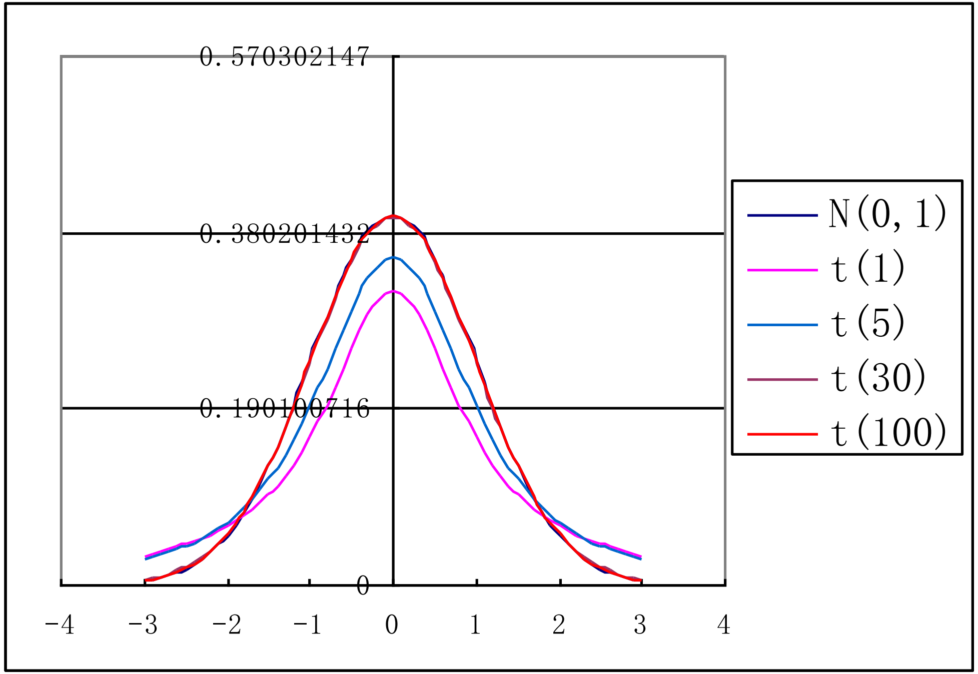
$$f(x;n)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}\ (-\infty<x<+\infty).$$

$$n=1\text{时，}\ f(x;1)=\frac{1}{\pi(1+x^2)},(-\infty<x<+\infty).$$

t 分布就是柯西分布.

当$n\to\infty$时，t分布的极限分布是$N(0,1)$.

t 分布的概率密度曲线



0.570302147
0.380201432
0.190100716
0
-4
-3
-2
-1
0
1
2
3
4
N(0, 1)
t(1)
t(5)
t(30)
t(100)

t 分布的上侧α分位数或临界值

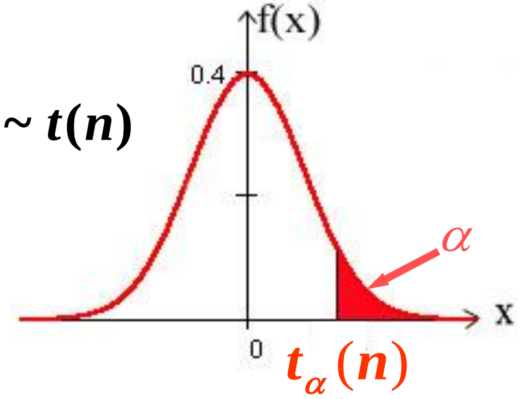
设$\alpha(0<\alpha<1)$,称满足等式

$$P(T \geq t_\alpha(n)) = \alpha$$

的点$t_\alpha(n)$为$t(n)$分布的上侧α分位数或临界值.

$t_\alpha(n)$的值可查t分布表获得.

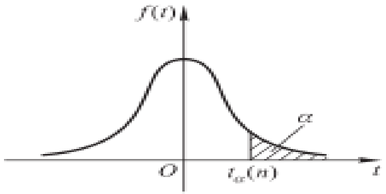
$T \sim t(n)$



$$P(T \geq t_\alpha(n)) = \int_{t_\alpha(n)}^{+\infty} f(x,n)dx = \alpha.$$

$$t_{0.025}(12) = 2.1788$$

附表4 t 分布表



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

$$t_{0.05}(9) = 1.8331,$$

n	α=0.25	0.10	0.05	0.025	0.01	0.005
1	1.000 0	3.077 7	6.313 8	12.706 2	31.820 7	63.657 4
2	0.816 5	1.885 6	2.920 0	4.302 7	6.964 6	9.924 8
3	0.764 9	1.637 7	2.353 4	3.182 4	4.540 7	5.840 9
4	0.740 7	1.533 2	2.131 8	2.776 4	3.746 9	4.604 1
5	0.726 7	1.475 9	2.015 0	2.570 6	3.364 9	4.032 2
6	0.717 6	1.439 8	1.943 2	2.446 9	3.142 7	3.707 4
7	0.711 1	1.414 9	1.894 6	2.364 6	2.998 0	3.499 5
8	0.706 4	1.396 8	1.859 5	2.306 0	2.896 5	3.355 4
9	0.702 7	1.383 0	1.833 1	2.262 2	2.821 4	3.249 8
10	0.699 8	1.372 2	1.812 5	2.228 1	2.763 8	3.169 3
11	0.697 4	1.363 4	1.795 9	2.201 0	2.718 1	3.105 8
12	0.695 5	1.356 2	1.782 3	2.178 8	2.681 0	3.054 5
13	0.693 8	1.350 2	1.770 9	2.160 4	2.650 3	3.012 3

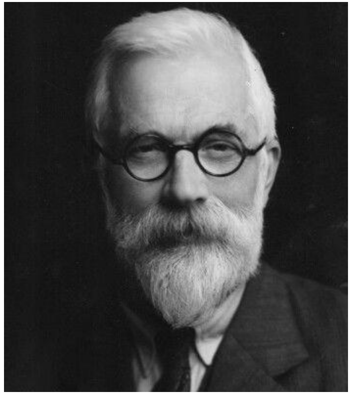
F分布

- 定义 设X,Y相互独立，且

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

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

服从第一自由度为n_1，第二自由度为n_2的F分布，记为$F\sim F(n_1,n_2)$.



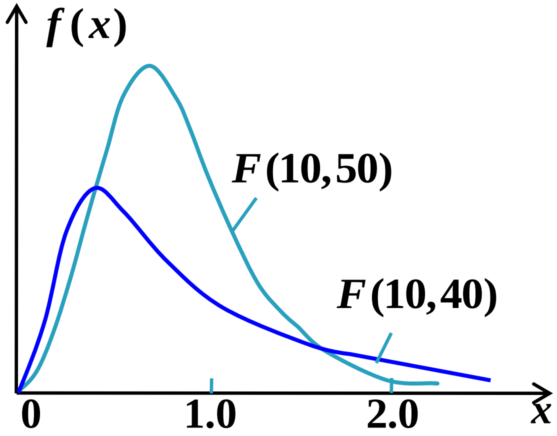
Ronald Fisher
1924提出了F分布.

- 若$F\sim F(n_1,n_2)$，则

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

F分布的概率密度函数

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

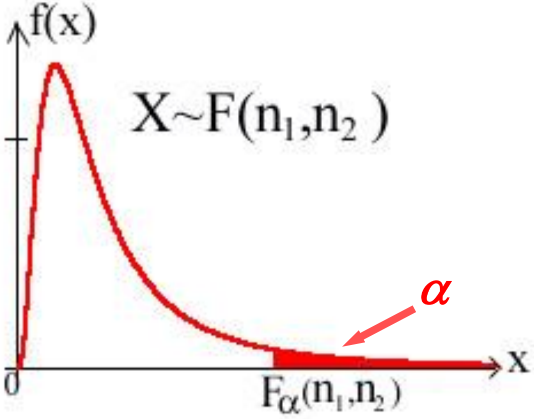


F分布的上侧α分位数或临界值

设 $\alpha(0<\alpha<1)$,称满足等式

$$P(F \geq F_\alpha(n_1,n_2)) = \alpha$$

的点$F_\alpha(n_1,n_2)$为$F(n_1,n_2)$分布的上侧α分位数或临界值.$F_\alpha(n_1,n_2)$的值可查F分布表获得.



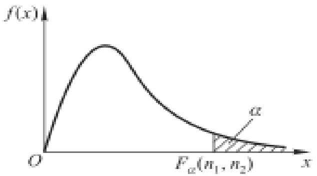
$$P(F \geq F_\alpha(n_1,n_2)) = \int_{F_\alpha(n_1,n_2)}^{+\infty} f(x,n)dx = \alpha.$$

$$F_{0.05}(9,12) = 2.8.$$

$\alpha=0.05$

n_2	n_1											
	1	2	3	4	5	6	7	8	9	10	12	15
1	161. 4	199. 5	215. 7	224. 6	230. 2	234. 0	236. 8	238. 9	240. 5	241. 9	243. 9	245. 9
2	18. 51	19. 00	19. 16	19. 25	19. 30	19. 33	19. 35	19. 37	19. 38	19. 40	19. 41	19. 43
3	10. 13	9. 55	9. 28	9. 12	9. 01	8. 94	8. 89	8. 85	8. 81	8. 79	8. 74	8. 70
4	7. 71	6. 94	6. 59	6. 39	6. 26	6. 16	6. 09	6. 04	6. 00	5. 96	5. 91	5. 86
5	6. 61	5. 79	5. 41	5. 19	5. 05	4. 95	4. 88	4. 82	4. 77	4. 74	4. 68	4. 62
6	5. 99	5. 14	4. 76	4. 53	4. 39	4. 28	4. 21	4. 15	4. 10	4. 06	4. 00	3. 94
7	5. 59	4. 74	4. 35	4. 12	3. 97	3. 87	3. 79	3. 73	3. 68	3. 64	3. 57	3. 51
8	5. 32	4. 46	4. 07	3. 84	3. 69	3. 58	3. 50	3. 44	3. 39	3. 35	3. 28	3. 22
9	5. 12	4. 26	3. 86	3. 63	3. 48	3. 37	3. 29	3. 23	3. 18	3. 14	3. 07	3. 01
10	4. 96	4. 10	3. 71	3. 48	3. 33	3. 22	3. 14	3. 07	3. 02	2. 98	2. 91	2. 85
11	4. 84	3. 98	3. 59	3. 36	3. 20	3. 09	3. 01	2. 95	2. 90	2. 85	2. 79	2. 72
12	4. 75	3. 89	3. 49	3. 26	3. 11	3. 00	2. 91	2. 85	2. 80	2. 75	2. 69	2. 62
13	4. 67	3. 81	3. 41	3. 18	3. 03	2. 92	2. 83	2. 77	2. 71	2. 67	2. 60	2. 53
14	4. 60	3. 74	3. 34	3. 11	2. 96	2. 85	2. 76	2. 70	2. 65	2. 60	2. 53	2. 46
15	4. 54	3. 68	3. 29	3. 06	2. 90	2. 79	2. 71	2. 64	2. 59	2. 54	2. 48	2. 40
16	4. 49	3. 63	3. 24	3. 01	2. 85	2. 74	2. 66	2. 59	2. 54	2. 49	2. 42	2. 35
17	4. 45	3. 59	3. 20	2. 96	2. 81	2. 70	2. 61	2. 55	2. 49	2. 45	2. 38	2. 31

附表5　F 分 布 表



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

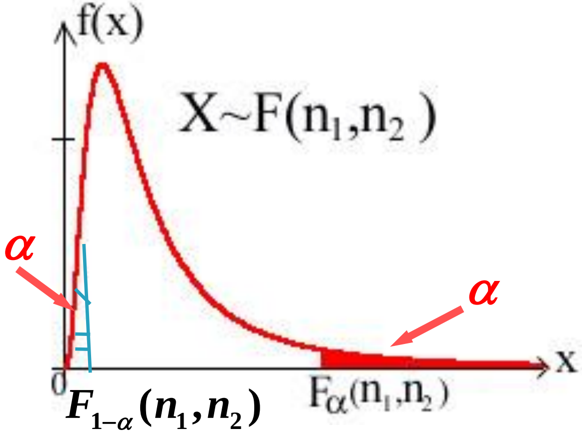
$\alpha=0.10$

$$F_{0.05}(9,12)=2.8.$$

$$F_{0.95}(12,9)=?$$

F分布的上侧α分位数或临界值

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



$$F_{0.05}(9,12)=2.8.$$

$$F_{0.95}(12,9)=\frac{1}{F_{0.05}(9,12)}$$

$$=\frac{1}{2.8}.$$

t分布与F分布的关系

例2 若随机变量$X\sim t(n)$, 则$X^2\sim F(1, n)$.

解

因为 $X\sim t(n)$, 所有$X = U\bigg/\sqrt{\dfrac{V}{n}}$,

这里$U\sim N(0,1), V\sim\chi^2(n)$, 且 U与V独立,

$$\text{故 } X^2=\left(\frac{U}{\sqrt{V/n}}\right)^2=\frac{U^2/1}{V/n}\sim F(1,n).$$

同理，若$X\sim t(n)$, 则$1/X^2\sim F(n, 1)$.

例1 设总体$X\sim N(0,4)$, 而X_1,X_2,X_3,X_4是总体X的样本，求下列随机变量的分布

（1） $T_1=(X_1^2+X_2^2)/4$; （2） $T_2=\dfrac{\sqrt{2}X_3}{\sqrt{X_1^2+X_2^2}}$;

（3） $T_3=\dfrac{2X_3^2}{X_1^2+X_2^2}$;

解（1） $X_1\sim N(0,4)\Rightarrow \dfrac{X_1}{2}\sim N(0,1)$,

同理$\dfrac{X_2}{2}\sim N(0,1)$,且二者相互独立,

故 $T_1 = \left(\dfrac{X_1}{2}\right)^2 + \left(\dfrac{X_2}{2}\right)^2 = \dfrac{X_1^2 + X_2^2}{4} \sim \chi^2(2).$

(2) $\dfrac{X_3}{2} \sim N(0,1),\ \dfrac{X_1^2 + X_2^2}{4} \sim \chi^2(2)$

且二者相互独立，故

$$T_2 = \frac{X_3/2}{\sqrt{\dfrac{X_1^2+X_2^2}{4}\Big/2}} = \frac{\sqrt{2}X_3}{\sqrt{X_1^2+X_2^2}} \sim t(2).$$

(3) $$\frac{X_3}{2} \sim N(0,1) \Rightarrow \left(\frac{X_3}{2}\right)^2 \sim \chi^2(1),\ \frac{X_1^2+X_2^2}{4} \sim \chi^2(2)$$

二者相互独立，故

$$T_3 = \frac{\left(\frac{X_3}{2}\right)^2 \Big/ 1}{\frac{X_1^2+X_2^2}{4} \Big/ 2} = \frac{2X_3^2}{X_1^2+X_2^2} \sim F(1,2).$$

谢　　谢！