§ 7.3 统计量的分布

确定统计量的分布—— 抽样分布, 是数理统计的基本问题之一．采用求随机向量的函数的分布的方法可得到抽样分布．由于样本容量一般不止 2 或 3 (甚至还可能是随机的)， 故计算往往很复杂，有时还需要特殊技巧或特殊工具.

由于正态总体是最常见的总体, 故本节介绍的几个抽样分布均对正态总体而言.

统计中常用分布

(1) 正态分布

若 $X_1, X_2, \cdots, X_n \quad \sim \quad N(\mu_i, \sigma_i^2)$

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

independent and identically distributed

特别地,

若 $X_1, X_2, \cdots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$

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

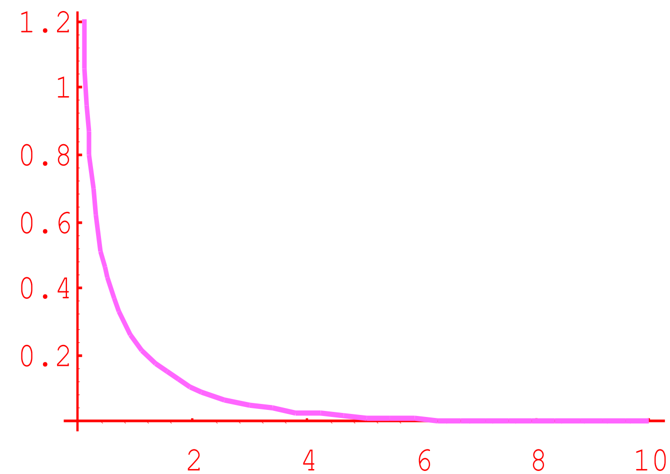
(2) $\chi^2(n)$ 分布（n为自由度）

定义 设$(X_1, X_2, \ldots, X_n)$相互独立，
且都服从标准正态分布$N(0,1)$，则称

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

$n=1$ 时，其密度函数为

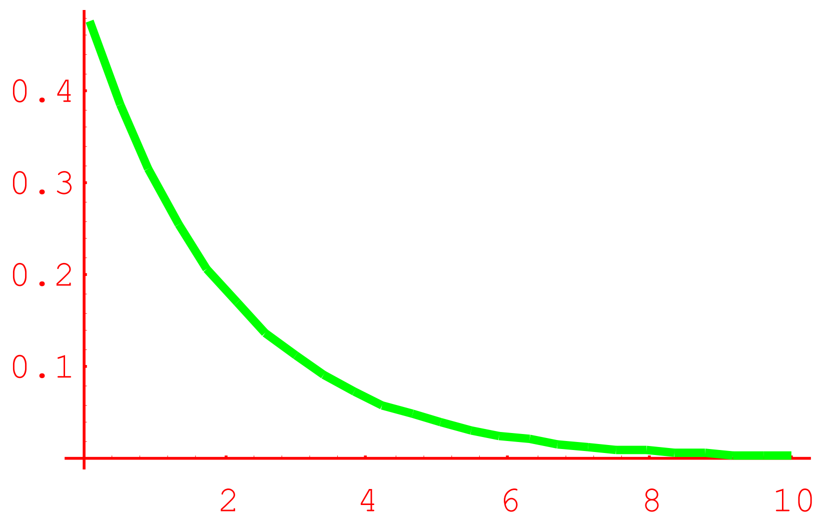
$$f(x)=\begin{cases} \dfrac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{x}{2}}, & x>0 \\ 0, & x\le 0 \end{cases}$$



$n=2$ 时,其密度函数为

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

为参数为1/2的指数分布.

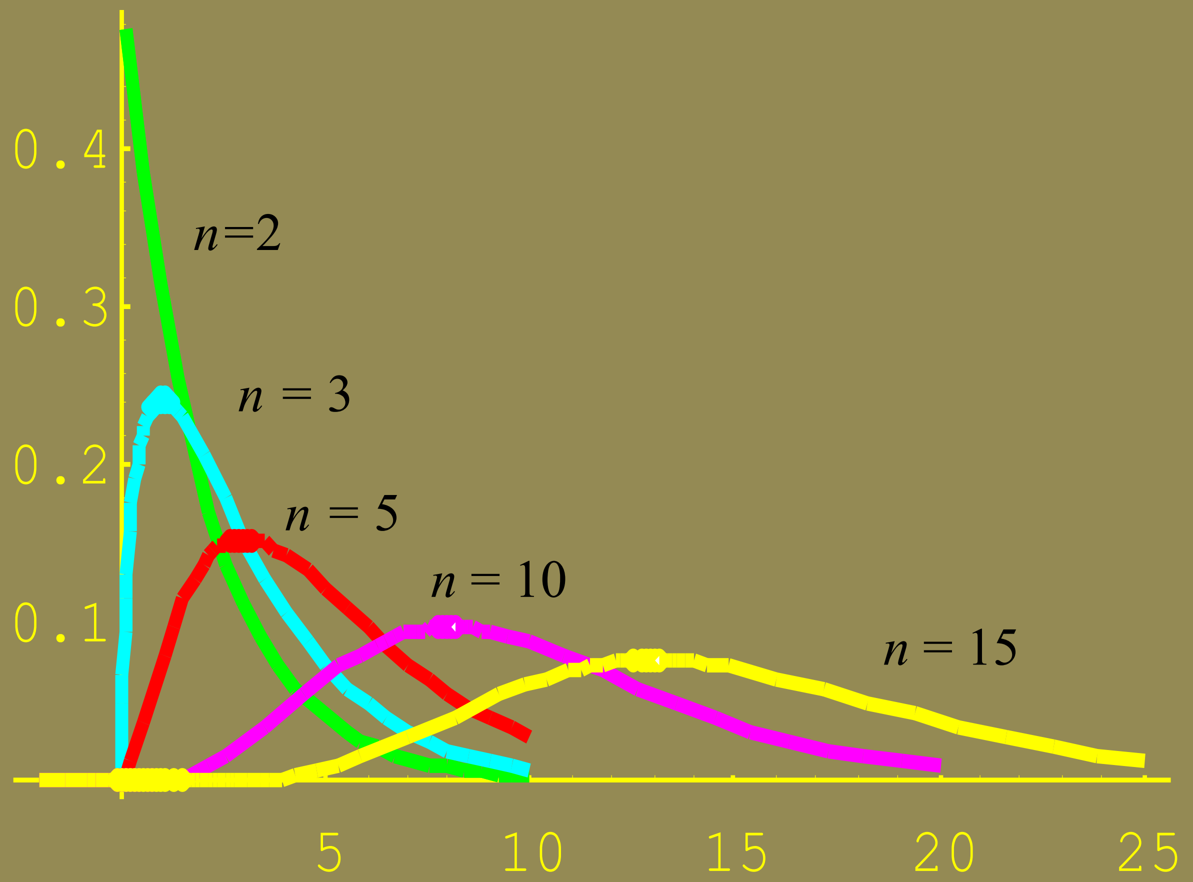


一般地, 自由度为 n 的$\chi^2(n)$的密度函数为

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

其中，$\Gamma(x)=\int_0^{+\infty}t^{x-1}e^{-t}dt$

在$x>0$时收敛，称为Γ函数.



0.4
0.3
0.2
0.1
5
10
15
20
25
n=2
n = 3
n = 5
n = 10
n = 15

$\chi^2(n)$ 分布的性质

1、$E\left(\chi^2(n)\right)=n, D\left(\chi^2(n)\right)=2n$

2、若 $X_1=\chi^2(n_1), X_2=\chi^2(n_2),$

X_1, X_2相互独立，则

$X_1+X_2 \sim \chi^2(n_1+n_2)$

3、 $\chi^2(n)$的α分位数有表可查。

n>45时 $\chi_\alpha^2(n) \approx \frac{1}{2}(z_\alpha+\sqrt{2n-1})^2$

证 1° 设 $\chi^2(n)=\sum_{i=1}^{n}X_i^2 \quad X_i \sim N(0,1) \ \ i=1,2,\cdots,n$

$X_1, X_2, \cdots, X_n$ 相互独立,

则 $E(X_i)=0,\ D(X_i)=1,\ E(X_i^2)=1$

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

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

$$D(X_i^2)=E(X_i^4)-E^2(X_i^2)=2$$

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

(3) t 分布 (Student 分布)

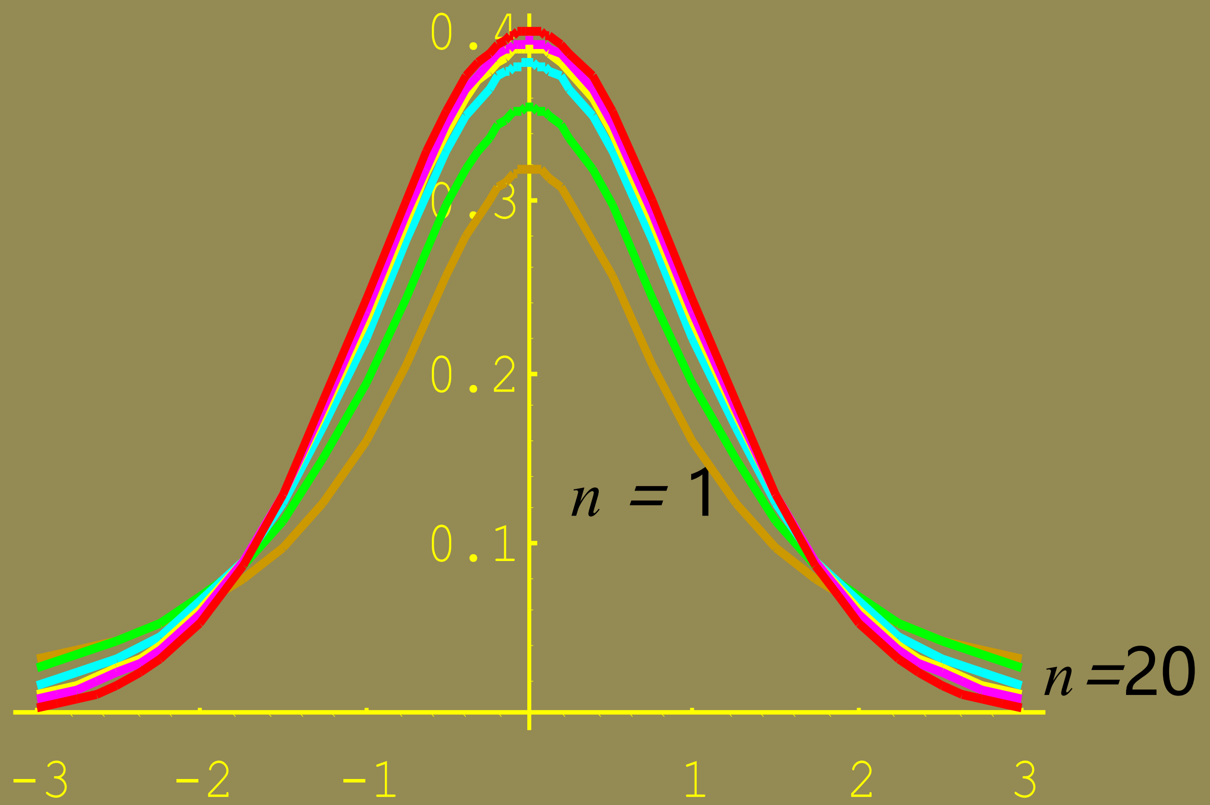
定义 设 $X \sim N(0,1)$，$Y \sim \chi^2(n)$, X, Y 相互独立,

$$T = \frac{X}{\sqrt{Y/n}}$$

则T所服从的分布称为自由度为 n 的T分布

其密度函数为

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



t 分布的图形(红色的是标准正态分布)

t 分布的性质

$1^\circ f_n(t)$是偶函数,

$$n \to \infty, f_n(t) \to \phi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

$2^\circ T$分布的α分位数 t_α有表可查

(4) F 分布

定义 设 $X \sim \chi^2(n)$, $Y \sim \chi^2(m)$, X , Y 相互独立,

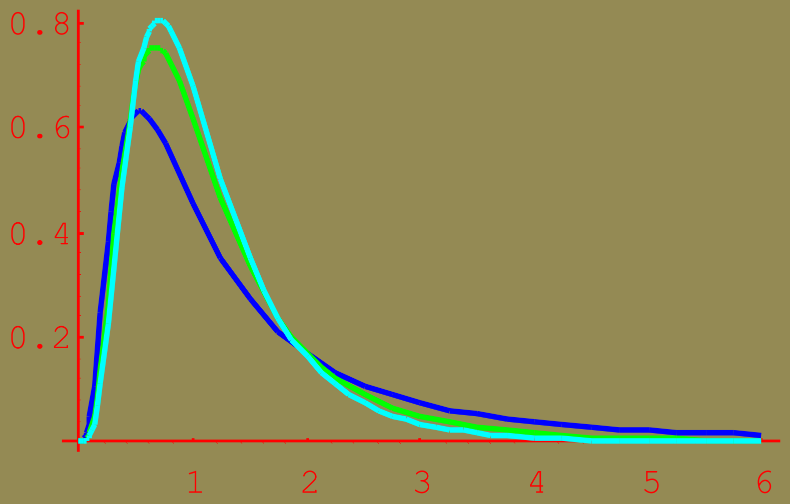
令
$$F = \frac{X/n}{Y/m}$$

则F 所服从的分布称为**第一自由度为n ，第二自由度为 m 的F 分布**

记为 $F \sim F(n,m)$

其密度函数为

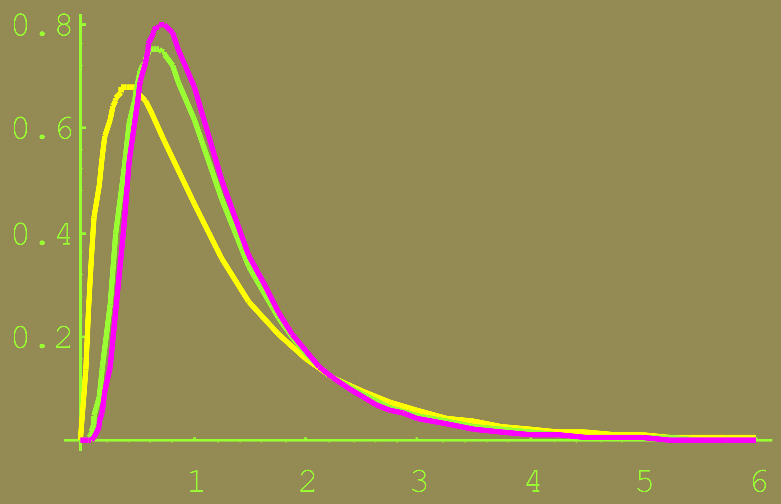
$$f(t,n,m)=\begin{cases}\dfrac{\Gamma\left(\dfrac{n+m}{2}\right)}{\Gamma\left(\dfrac{n}{2}\right)\Gamma\left(\dfrac{m}{2}\right)}\left(\dfrac{n}{m}\right)^{\frac{n}{2}}t^{\frac{n}{2}-1}\left(1+\dfrac{n}{m}t\right)^{-\frac{n+m}{2}} & t>0\\ 0, & t\le 0\end{cases}$$



$m = 10, n = 4$

$m = 10, n = 10$

$m = 10, n = 15$



$m = 4, n = 10$

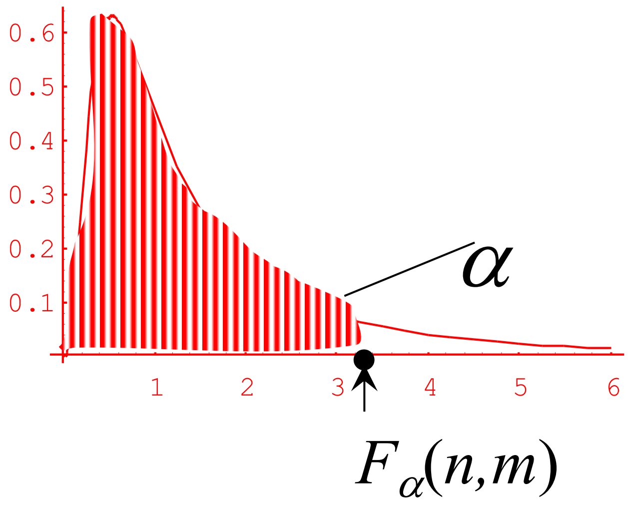
$m = 10, n = 10$

$m = 15, n = 10$

F 分布的性质

1、$F(n,m)$的α分位数$F_\alpha(n,m)$有表可查:

$$P(F \le F_\alpha(n,m)) = \alpha$$



2、若$F\sim F(n, m)$,则 $\dfrac{1}{F} \sim F(m,n)$

事实上，若 $F \sim F(n,m)$

则可设 $F = \dfrac{X/n}{Y/m}$

故有 $\dfrac{1}{F} = \dfrac{Y/m}{X/n} \sim F(m,n)$

从而，在查表时可能用到结论：$F_{1-\alpha}(n,m)=\dfrac{1}{F_{\alpha}(m,n)}$

$$1-\alpha=P(F\le F_{1-\alpha}(n,m))\ =P(\frac{1}{F}>\frac{1}{F_{1-\alpha}(n,m)})$$

$$=1-P(\frac{1}{F}\le\frac{1}{F_{1-\alpha}(n,m)})$$

即 $P(\dfrac{1}{F}\le\dfrac{1}{F_{1-\alpha}(n,m)})=\alpha$ 　而 $\dfrac{1}{F}\sim F(m,n)$

即 $F_{\alpha}(m,n)=\dfrac{1}{F_{1-\alpha}(n,m)}$ 　$F_{1-\alpha}(n,m)=\dfrac{1}{F_{\alpha}(m,n)}$

例如 $F_{0.95}(5,4)=6.26$

故 $F_{0.05}(4,5)=\dfrac{1}{F_{0.95}(5,4)}=0.159$