三个常用的抽样分布

统计量的分布称为抽样分布.

1. χ^2分布 (Chi-square Distribution)

由海尔墨特(**Hermert**)于**1875**提出，是统计学中的一个非常有用的著名分布。 现代统计学的奠基人之一的卡尔·皮尔逊(**K. Pearson**)于**1900**年提出了卡方检验。

作用：主要用于拟合优度检验，独立性检验以及对总体方差的估计和检验。

设 $X_1, X_2, \cdots, X_n$ 是来自总体 $N(0,1)$ 的样本，则称统计量 $\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2$ 服从自由度为 n 的 χ^2 分布，记为 $\chi^2 \sim \chi^2(n)$.

自由度:

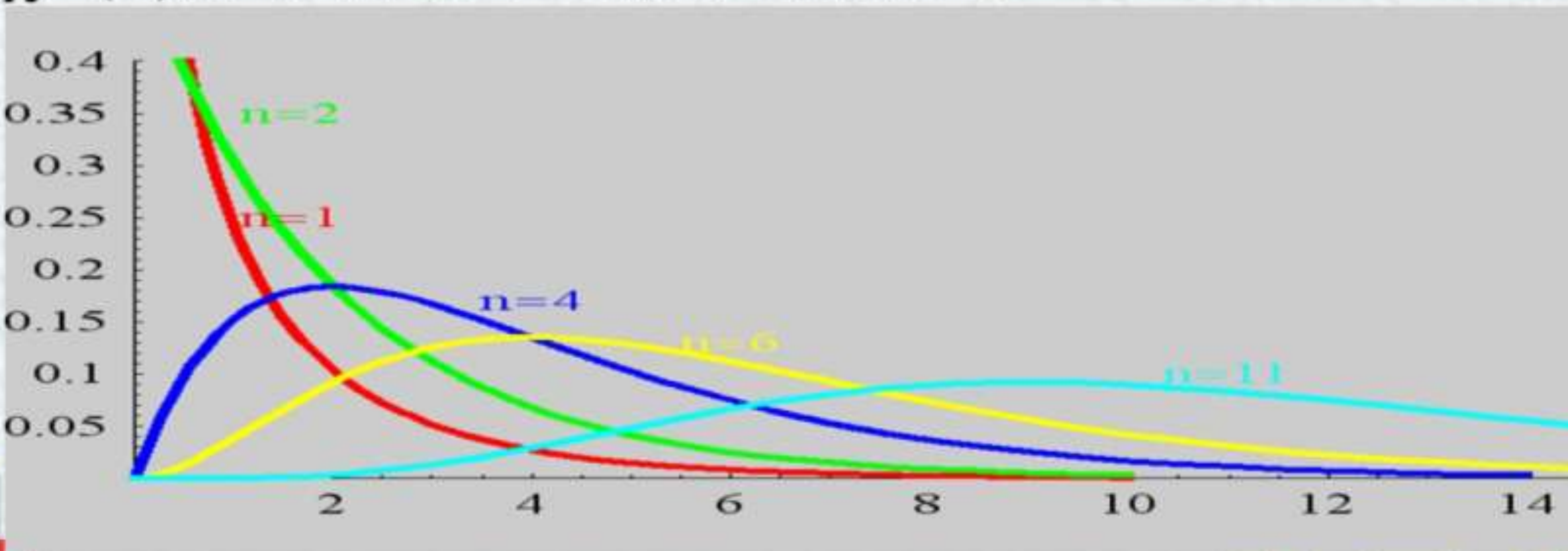
指 $\chi^2 = X_1^2 + X_2^2 + \cdots + X_n^2$ 中右端包含独立变量的个数.

$\chi^2(n)$分布的概率密度为

书P133 (3.53)

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

$\chi^2(n)$分布的概率密度曲线如图.



例1：设总体X $\sim N\left(\mu,\sigma^2\right)$, μ,σ^2 已知。$\left(X_1,X_2,\cdots,X_n\right)$是取自总体X的样本

求(1)统计量 $\chi^2=\dfrac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$ 的分布；

（2）设n=5, 若a$(X_1-X_2)^2+b(2X_3-X_4-X_5)^2\sim\chi^2(k)$,

则a, b, k各为多少？

解：(1)作变换 $Y_i=\dfrac{X_i-\mu}{\sigma}\quad i=1,2,\cdots,n$

显然$Y_1,Y_2,\cdots,Y_n$相互独立，且$Y_i\sim N\left(0,1\right)\quad i=1,2,\cdots,n$

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

(2) $X_1-X_2\sim N(0,2\sigma^2)$, $\dfrac{(X_1-X_2)^2}{2\sigma^2}\sim\chi^2(1)$

$2X_3-X_4-X_5\sim N(0,6\sigma^2)$, $\dfrac{(2X_3-X_4-X_5)^2}{6\sigma^2}\sim\chi^2(1)$

X_1-X_2与$2X_3-X_4-X_5$相互独立，

故$\dfrac{(X_1-X_2)^2}{2\sigma^2}+\dfrac{(2X_3-X_4-X_5)^2}{6\sigma^2}\sim\chi^2(2)$

$a=\dfrac{1}{2\sigma^2}$,

$b=\dfrac{1}{6\sigma^2}$,

$k=2.$

重要结论

定理 设 $(X_1, X_2, \cdots, X_m)$ 和 $(Y_1, Y_2, \cdots, Y_n)$ 相互独立,则 $X_i(1,2,\cdots,m)$ 和 $Y_j(j=1,2,\cdots,n)$ 相互独立.又若 h, g 是连续函数,则 $h(X_1, X_2, \cdots, X_m)$ 和 $g(Y_1, Y_2, \cdots, Y_n)$ 相互独立.

χ^2 分布的性质

性质1 (χ^2 分布的可加性)

设 $X \sim \chi^2(n_1)$, $Y \sim \chi^2(n_2)$, 并且 X, Y 独立, 则 $X+Y \sim \chi^2(n_1+n_2)$.

(此性质可以推广到多个随机变量的情形.)

若 $X_i \sim \chi^2(n_i)$, 并且 X_i $(i=1,2,\cdots,m)$ 相互独立, 则 $\sum_{i=1}^{m} X_i \sim \chi^2(n_1+n_2+\cdots+n_m)$.

性质2 (χ^2分布的数学期望和方差)

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

证明　因为 $X_i \sim N(0,1)$, 所以 $E(X_i^2) = D(X_i) = 1$,

$D(X_i^2) = E(X_i^4) - [E(X_i^2)]^2 = 3 - 1 = 2, \ i = 1, 2, \cdots, n.$

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

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

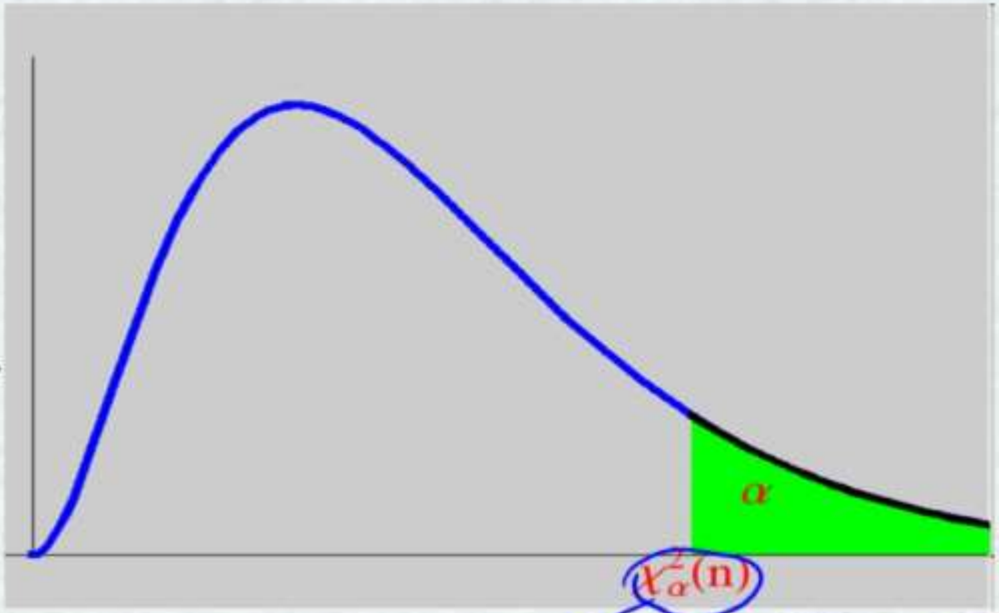
χ^2 分布的分位点

对于给定的正数α, $0<\alpha<1$, 称满足条件

$$P\{\chi^2 > \chi^2_{1-\alpha}(n)\} = \alpha$$

的点 $\chi^2_{1-\alpha}(n)$ 为 $\chi^2(n)$ 分布的上 α 分位点.

对于不同的α, n, 可以通过查表求得上α分位点的值.



$\chi^2_{1-\alpha}(n)$

例1 $\chi^2_{0.025}(10) = \mathbf{3.247}$,

$\chi^2_{0.9}(25) = \mathbf{34.382}$

例2 设 $\chi^2 \sim \chi^2(8)$，试确定 λ_1, λ_2 的值，使之满足：

$$P(\chi^2 > \lambda_1) = 0.05, \quad P(\chi^2 < \lambda_2) = 0.05.$$

解 $\lambda_1 = \chi^2_{0.95}(8) = \mathbf{15.507}$,

$\lambda_2 = \chi^2_{0.05}(8) = \mathbf{2.733}$,

2. t 分布

提出：其推导由威廉·戈塞特（**Gosset**）于1908年首先发表，当时他还在都柏林的健力士酿酒厂工作。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。之后t检验以及相关理论经由罗纳德·费希尔（ **R.A.Fisher** ）的工作发扬光大，而正是他将此分布称为学生氏(**Student**)分布。

应用：用于根据小样本来估计呈正态分布且方差未知的总体的均值。

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

则称随机变量 $t = \dfrac{X}{\sqrt{Y/n}}$ 服从自由度为 n 的 t

分布, 记为 $t \sim t(n)$.

$t(n)$ 分布的概率密度函数为

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

例

设X_1, X_2, X_3, X_4, X_5为来自标准正态总体$X \sim N(0,1)$的一组样本,试求下列随机变量的分布.

(1) $\dfrac{2X_1}{\sqrt{X_2^2+X_3^2+X_4^2+X_5^2}}$; (2) $\dfrac{X_2}{|X_1|}$.

解

由题意知, $X_i \sim N(0,1), i=1,2,3,4,5$, 且相互独立, 故由$\chi^2$分布的定义知

$$X_2^2+X_3^2+X_4^2+X_5^2 \sim \chi^2(4), \qquad X_1^2 \sim \chi^2(1).$$

因此, 再由t分布的定义可得

(1) $$\frac{2X_1}{\sqrt{X_2^2+X_3^2+X_4^2+X_5^2}} = \frac{X_1}{\sqrt{(X_2^2+X_3^2+X_4^2+X_5^2)/4}} \sim t(4);$$

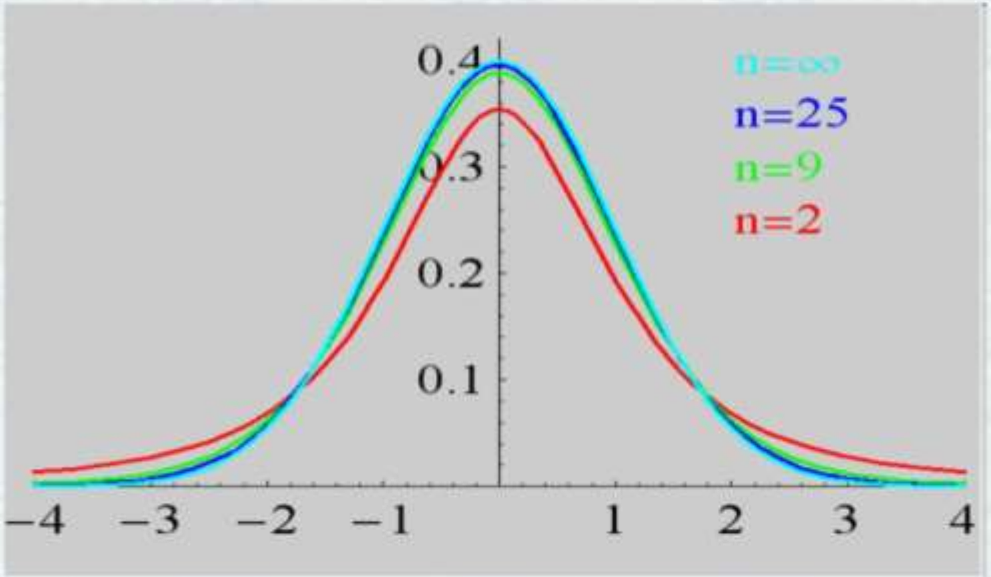
(2) $$\frac{X_2}{|X_1|} = \frac{X_2}{\sqrt{X_1^2/1}} \sim t(1).$$

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$\lim\limits_{n\to\infty} h(t) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$,

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.

t 分布的分位点

对于给定的 α, $0<\alpha<1$, 称满足条件

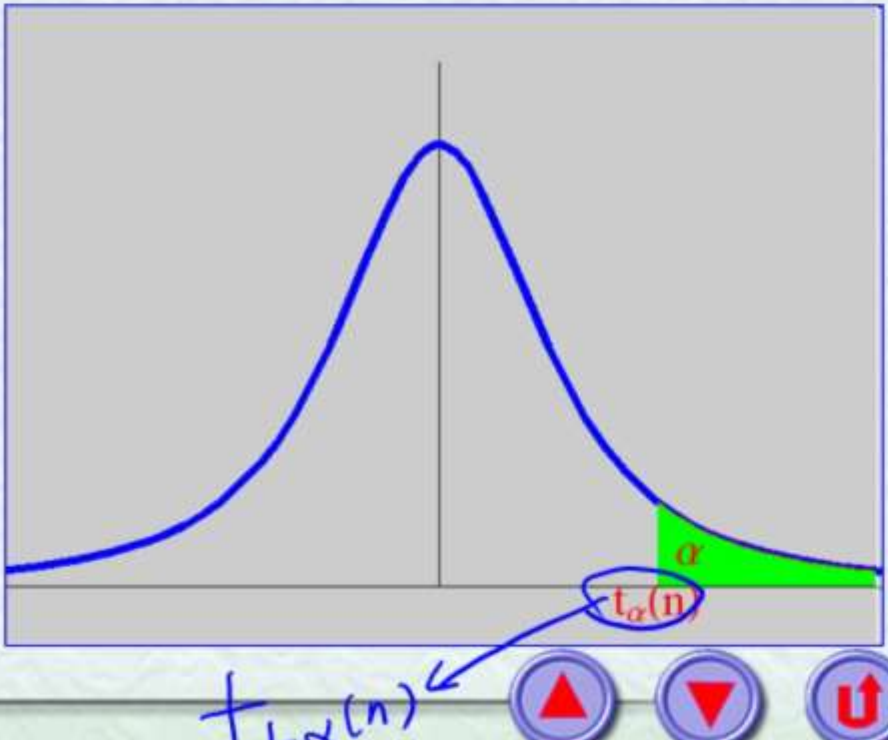
$$P\{t>t_{1-\alpha}(n)\}=\alpha$$

的点 $t_{1-\alpha}(n)$ 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

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



例3 $t_{0.95}(10) = \mathbf{1.8125},$

$t_{0.025}(15) = \mathbf{-2.1315.}$

例4 设 $T \sim t(14)$，试确定 λ_1, λ_2 的值，使之满足：

$$P(T > \lambda_1) = 0.05, \quad P(|T| < \lambda_2) = 0.95.$$

解 $\lambda_1 = t_{0.95}(14) = \mathbf{1.7613},$

$$P(T > \lambda_2) = \frac{0.05}{2} = 0.025$$

$\Rightarrow \lambda_2 = t_{0.975}(14) = \mathbf{2.1448},$

3. *F*分布

提出：F分布是1924年英国统计学家R.A.Fisher提出，并以其姓氏的第一个字母命名的。

应用：在方差分析、协方差分析及回归方程的显著性检验中都有着重要的地位。

设$U \sim \chi^2(n_1)$, $V \sim \chi^2(n_2)$, 且U, V 独立, 则称随机变量 $F = \dfrac{U/n_1}{V/n_2}$ 服从自由度为(n_1, n_2)的 F 分布, 记为 $F \sim F(n_1, n_2)$.

它是一种非对称分布，有两个自由度，且位置不可互换。

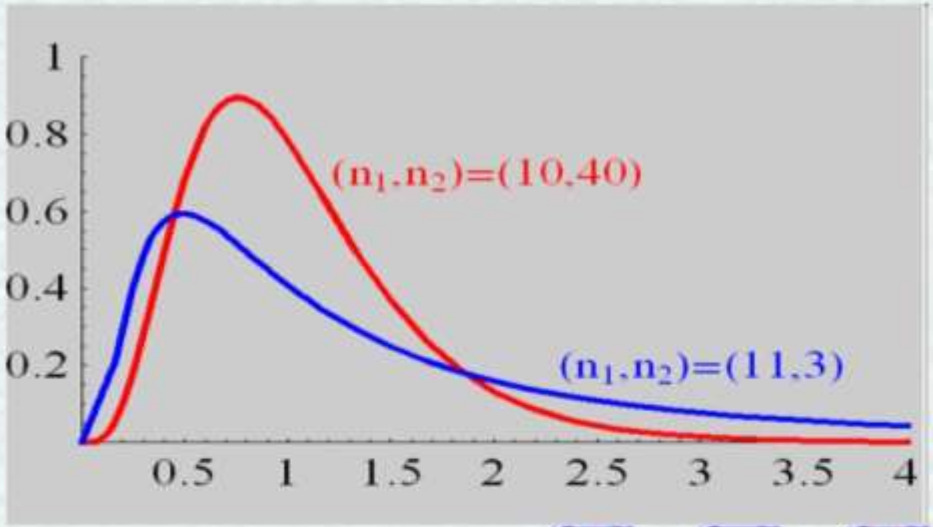
根据定义可知,

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

$F(n_1, n_2)$分布的概率密度为

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

F分布的概率密度曲线如图



例5 （1）设 X_1, X_2, X_3, X_4 是来自标准正态总体 $X \sim N(0,1)$的一组样本，试求 $\dfrac{X_1^2 + X_2^2}{X_3^2 + X_4^2}$ 分布；

（2）设$T \sim t(n)$, 试求T^2的分布.

解

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

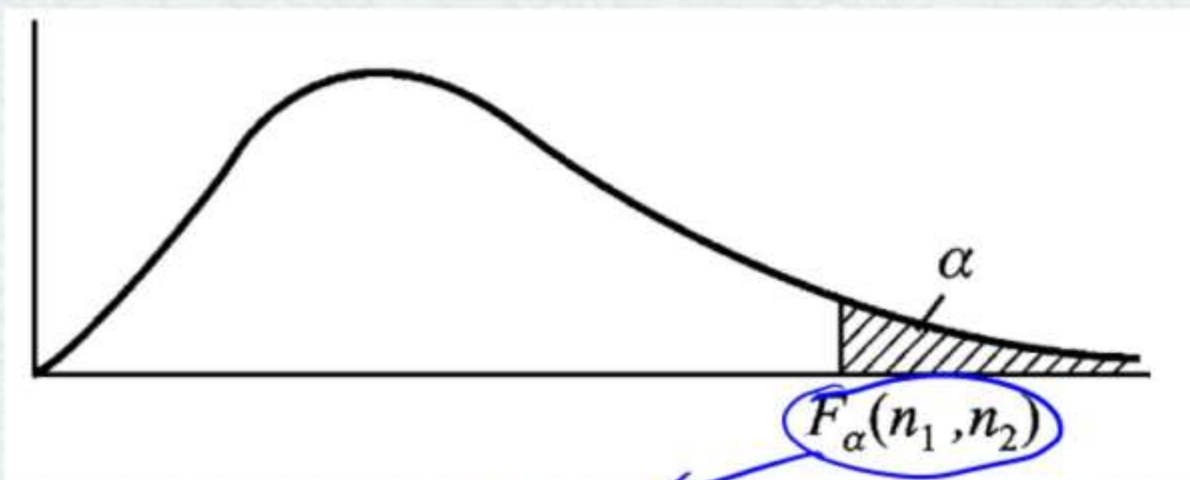
(2) $$T^2 = \left(\frac{X}{\sqrt{Y/n}}\right)^2 = \frac{X^2}{Y/n} \sim F(1,n)$$

F 分布的分位点

对于给定的α, $0<\alpha<1$, 称满足条件

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

的点 $F_{1-\alpha}(n_1, n_2)$ 为 $F(n_1, n_2)$ 分布的上 α 分位点.



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

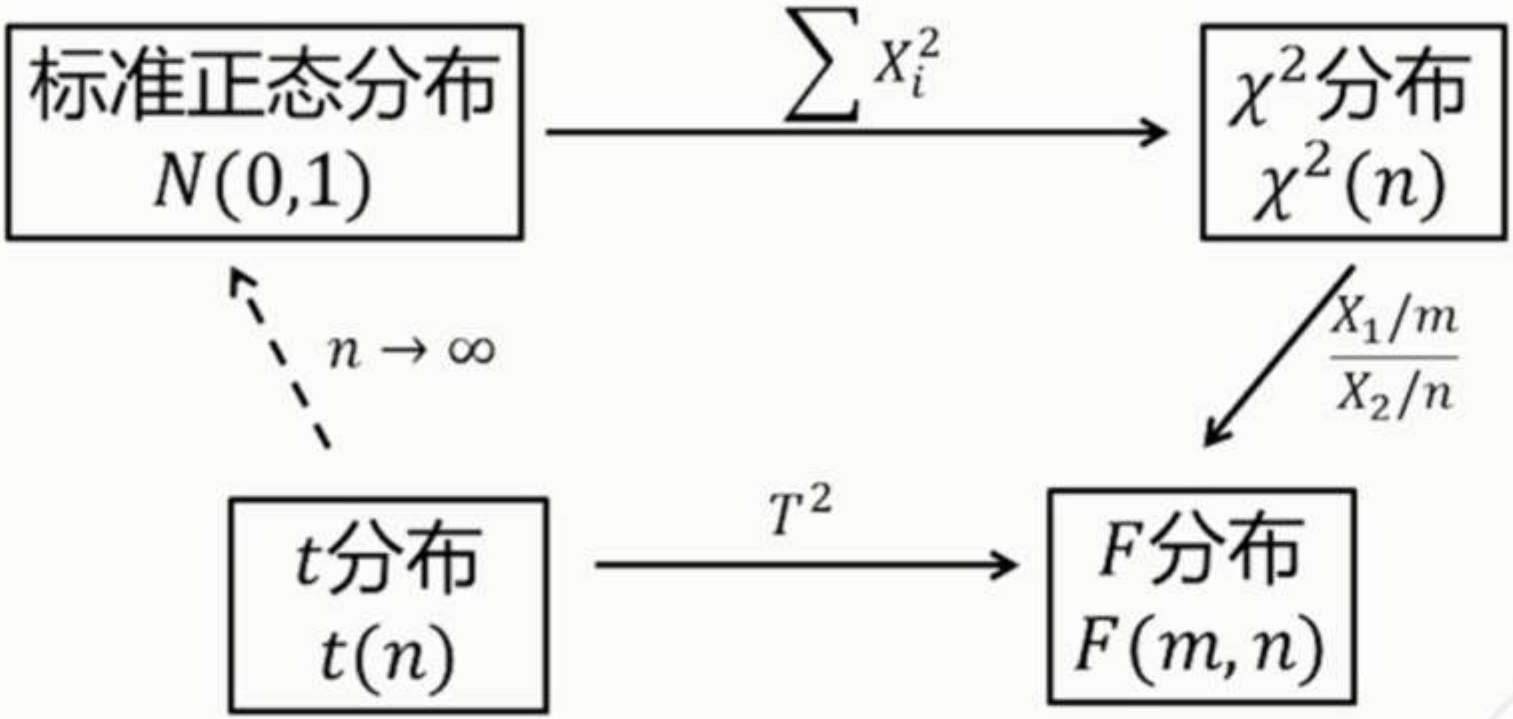
例6 $F_{0.9}(7,8) = 2.62,$

$F_{0.95}(14,30) = 2.04.$

F 分布的上 α 分位点具有如下性质:

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

例7 $F_{0.05}(12,9) = \dfrac{1}{F_{0.95}(9,12)} = \dfrac{1}{2.80} = 0.357.$


标准正态分布
$N(0,1)$
$\sum X_i^2$
χ^2分布
$\chi^2(n)$
$n \to \infty$
$\frac{X_1/m}{X_2/n}$
t分布
$t(n)$
T^2
F分布
$F(m,n)$