概率论与数理统计

第六章　样本及抽样分布

§3　抽样分布

- ◆统计量与经验分布函数
- ◆统计学三大抽样分布
- ◆几个重要的抽样分布定理
- ◆小结

回顾

正态分布

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

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

特别地,

若 $X_1, X_2, \cdots, X_n \sim X_i \sim N(\mu, \sigma^2)$

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

标准正态分布的上 α 分位点

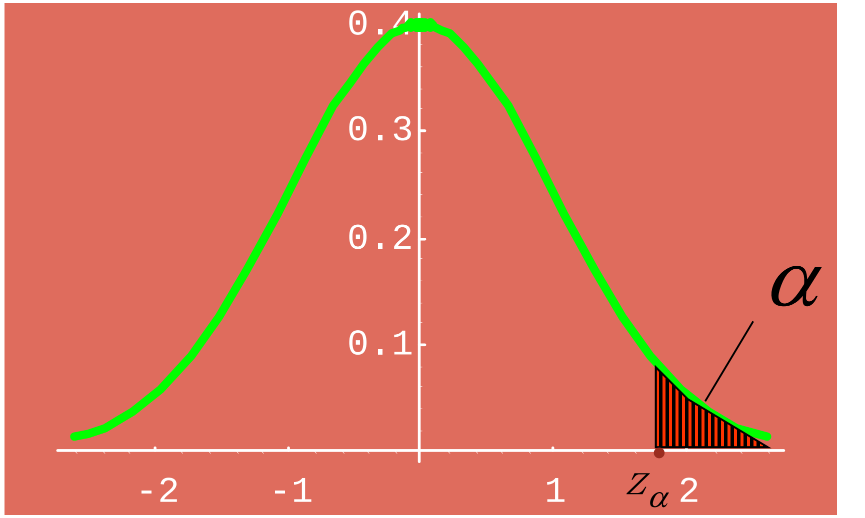
定义

若 $P(X > z_\alpha) = \alpha$，则称z_α为标准正态分布的上α 分位点.

若 $P(|X| > z_{\alpha/2}) = \alpha$，则称 $z_{\frac{\alpha}{2}}$为标准正态分布的双侧 α 分位点.

标准正态分布的上α分位点图形

$$P(X > z_\alpha) = \alpha$$



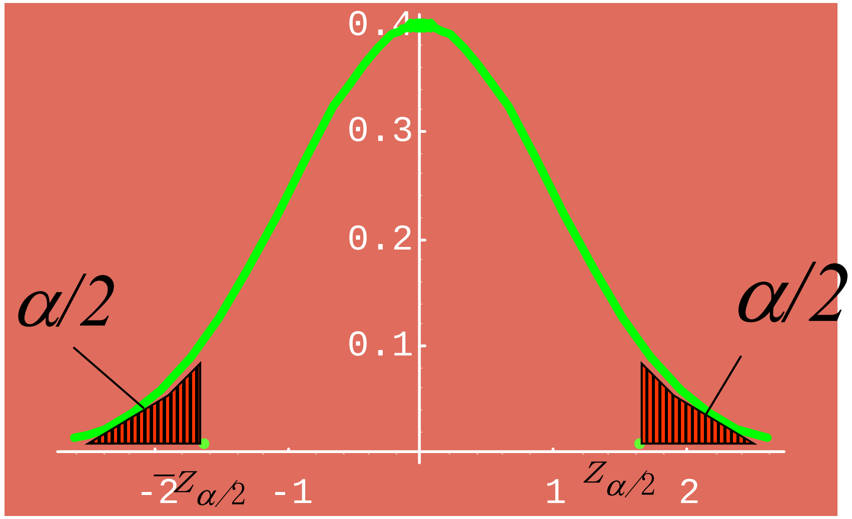
常用数字

$$z_{0.05} = 1.645$$

$$z_{0.025} = 1.96$$

$$z_{0.005} = 2.575$$

$$P(|X| > z_{\alpha/2}) = \alpha$$



二、常见分布

统计量的分布称为抽样分布.

1. χ^2分布

χ^2分布是由正态分布派生出来的一种分布.

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$\chi^2 \sim \chi^2(n)$.

自由度是指上式右端包含的独立变量的个数.

$\chi^2(n)$分布的概率密度为

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

证明　因为 $\chi^2(1)$ 分布即为 $\Gamma\left(\dfrac{1}{2},2\right)$ 分布,

又因为 $X_i\sim N(0,1)$, 由定义 $X_i^2\sim\chi^2(1)$,

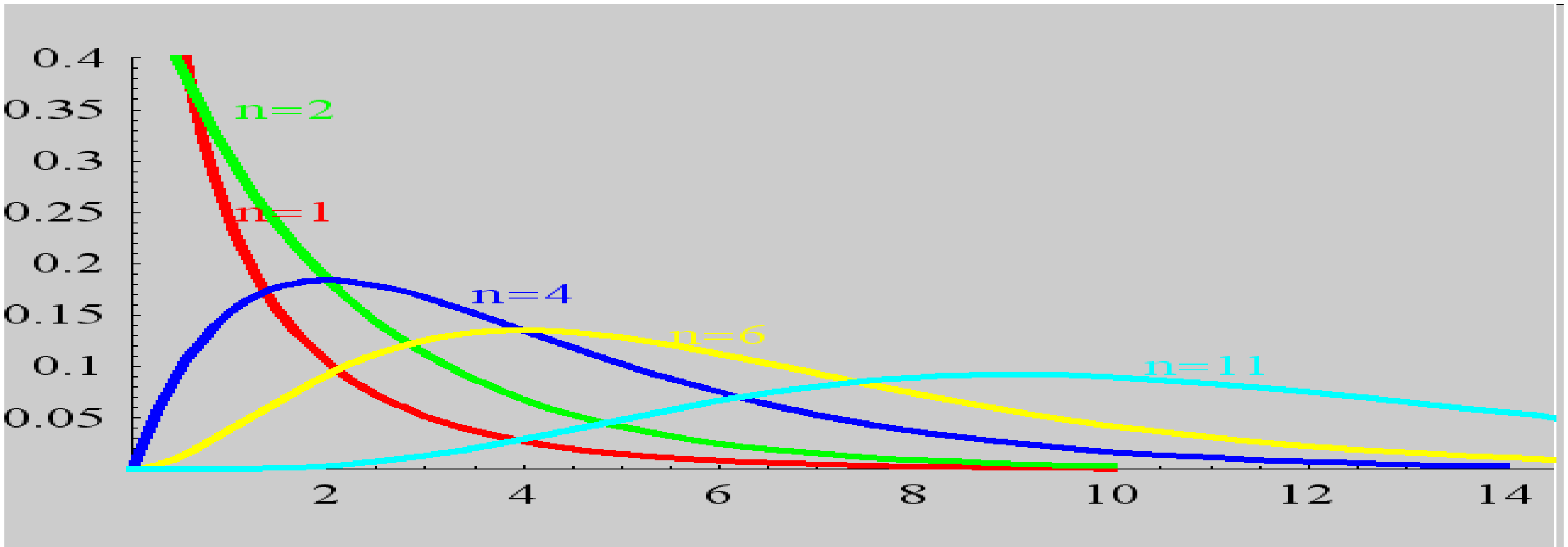
即 $X_i^2\sim\Gamma\left(\dfrac{1}{2},2\right)$,　$i=1,2,\cdots,n.$

因为 $X_1, X_2, \cdots, X_n$ 相互独立,

所以 $X_1^2, X_2^2, \cdots, X_n^2$ 也相互独立,

根据Γ分布的可加性知 $\chi^2 = \sum_{i=1}^{n} X_i^2 \sim \Gamma\left(\frac{n}{2}, 2\right)$.

$\chi^2(n)$分布的概率密度曲线如图.



注意： 其中伽玛函数 $\Gamma(x)$通过积分

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

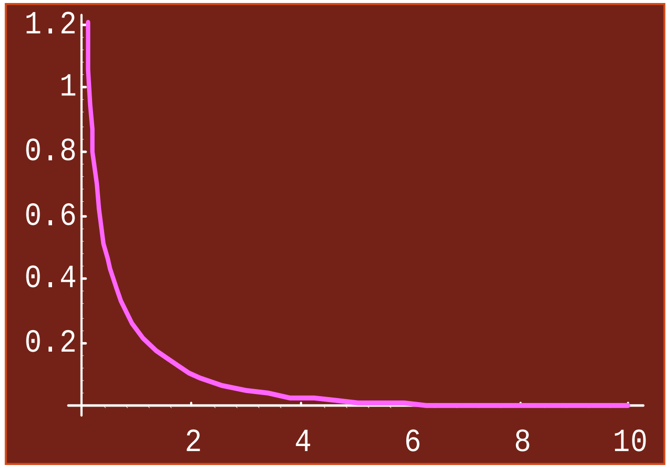
在$\boldsymbol{x>0}$时收敛，称为Γ函数，其具有以下性质

$$\Gamma(x+1)=x\Gamma(x), \quad \Gamma(1)=1, \ \Gamma(1/2)=\sqrt{\pi}$$

$$\Gamma(n+1)=n! \ (n\in N)$$

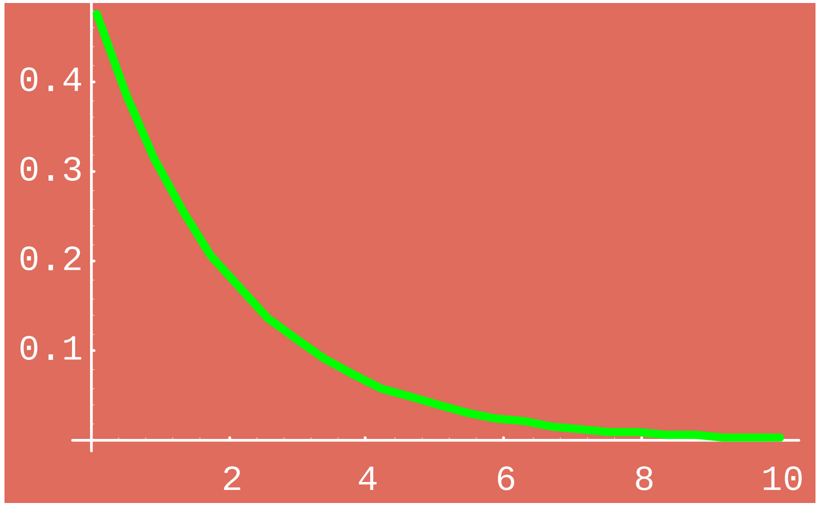
$n=1$ 时,其概率密度为

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



$n=2$ 时,其概率密度为

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



为参数为2的指数分布.

χ^2 分布的性质

性质1 (χ^2 分布的可加性)

设 $\chi_1^2 \sim \chi^2(n_1)$, $\chi_2^2 \sim \chi^2(n_2)$, 并且 χ_1^2, χ_2^2 独立, 则 $\chi_1^2 + \chi_2^2 \sim \chi^2(n_1 + n_2)$.

(此性质可以推广到多个随机变量的情形.)

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

性质2 (χ^2分布的数学期望和方差)

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

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

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

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

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

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

性质3 设 $X_1, X_2, \cdots, X_n$ 相互独立, 都服从正态分布 $N(\mu, \sigma^2)$, 则 $\chi^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi^2(n)$

性质4 若 $\chi^2 \sim \chi^2(n)$, 则当 n 充分大时, $\frac{\chi^2 - n}{\sqrt{2n}}$ 的分布近似正态分布 $N(0,1)$.

(应用中心极限定理可得)

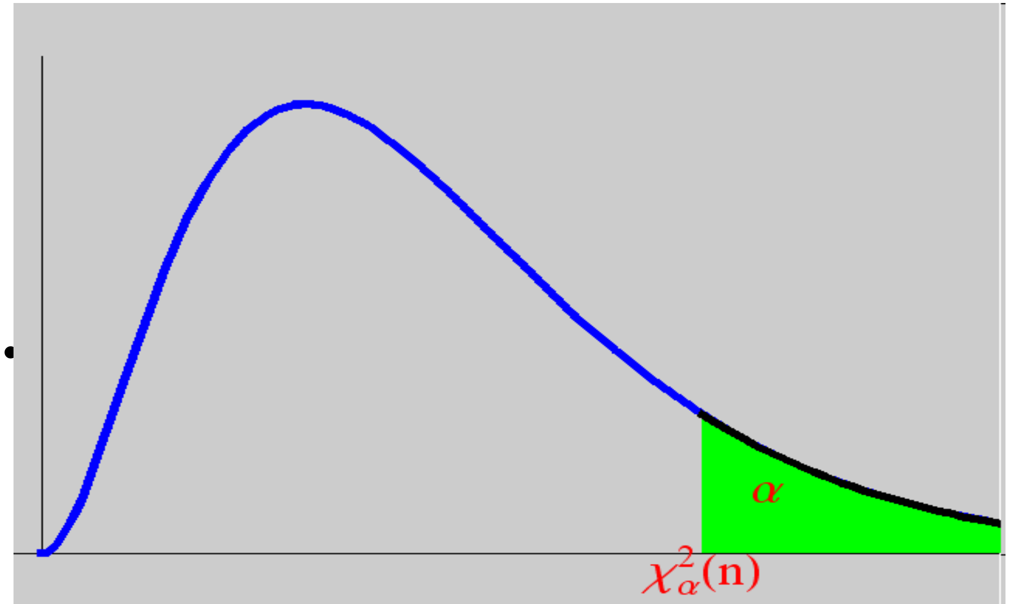
χ^2 分布的分位点

对于给定的正数 α，$0<\alpha<1$，称满足条件

$$P\{\chi^2>\chi^2_\alpha(n)\}=\int_{\chi^2_\alpha(n)}^{\infty}f(y)\mathrm{d}y=\alpha$$

的点 $\chi^2_\alpha(n)$ 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.



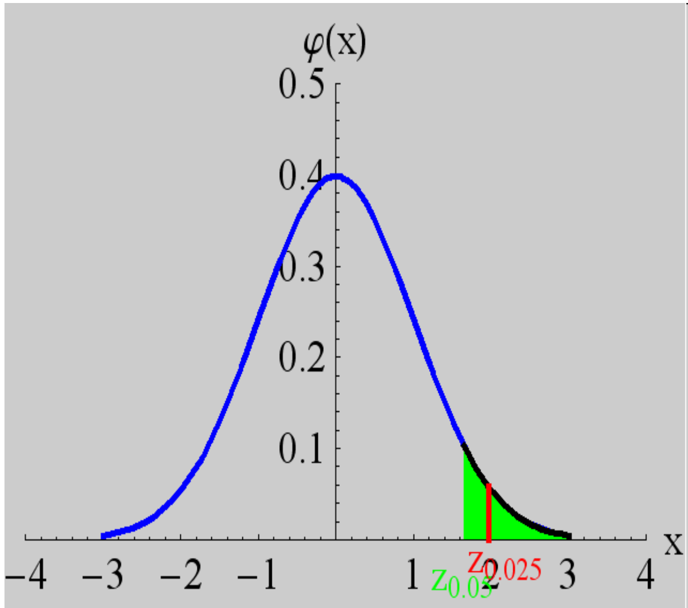
例1 设 X 服从标准正态分布 $N(0,1)$, $N(0,1)$ 的上 α 分位点 z_α 满足 $P\{X > z_\alpha\} = \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{z_\alpha}^{+\infty} e^{-\frac{x^2}{2}}\,dx = \alpha$, 求 z_α 的值, 可通过查表完成.

$z_{0.05} = 1.645$,

$z_{0.025} = 1.96$,

根据正态分布的对称性知

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



例2 设 $Z \sim \chi^2(n)$, $\chi^2(n)$ 的上 α 分位点满足

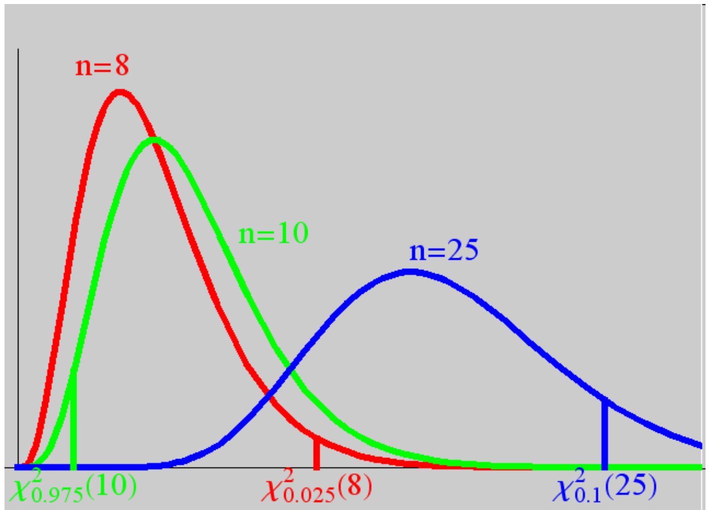
$$P\{Z > \chi_\alpha^2(n)\} = \int_{\chi_\alpha^2(n)}^{+\infty} \chi^2(y; n)\mathrm{d}y = \alpha,$$

求$\chi_\alpha^2(n)$的值，可通过查表完成.

$\chi_{0.025}^2(8) = 17.535,$

$\chi_{0.975}^2(10) = 3.247,$

$\chi_{0.1}^2(25) = 34.382.$



课本附表只详列到n=40为止.

费希尔(**R.A.Fisher**)证明:

当 n 充分大时, $\chi_{\alpha}^{2}(n) \approx \frac{1}{2}(z_{\alpha} + \sqrt{2n-1})^{2}.$

其中 z_{α} 是标准正态分布的上 α 分位点.

利用上面公式,

可以求得 $n > 40$ 时, 上 α 分位点的近似值.

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

而查详表可得 $\chi_{0.05}^{2}(50) = 67.505$.

2. t 分布

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 分布又称学生氏(Student)分布.

$t(n)$ 分布的概率密度函数为

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

注：具有自由度为n的t分布$t \sim t(n)$,其数学期望

与方差为：$E(t)=0, D(t)=n/(n-2) \qquad (n>2)$

t 分布的概率密度曲线如图

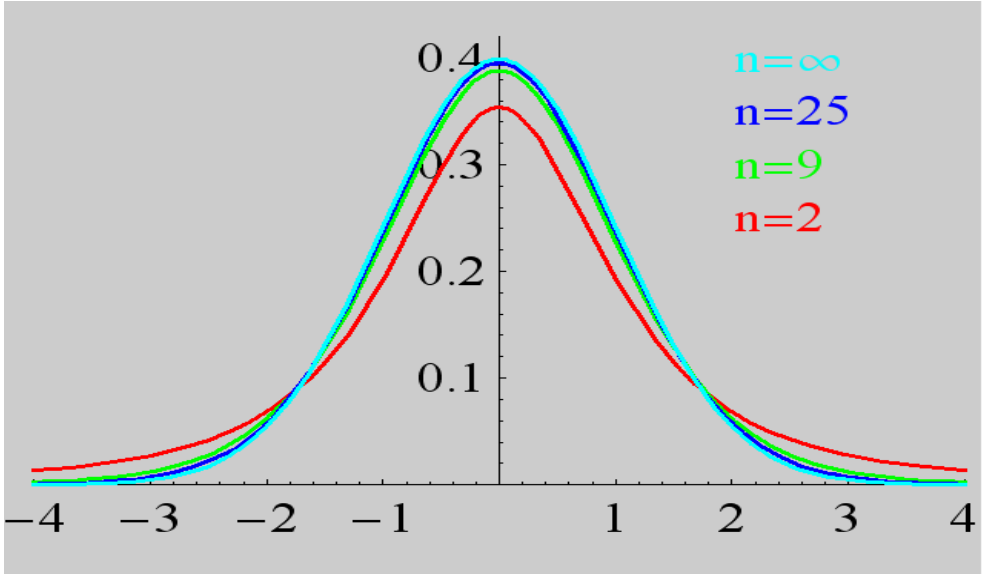
显然图形是关于

$t=0$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$\lim\limits_{n\to\infty} h(t)=\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$,

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.

注：$f_n(t)$是偶函数

t 分布的分位点

对于给定的 α，$0<\alpha<1$，称满足条件

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

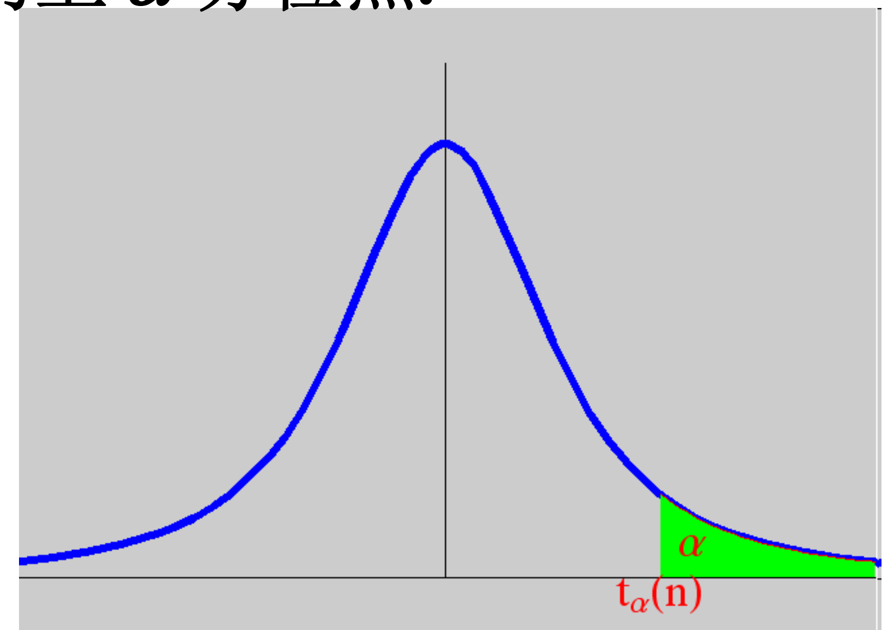
的点 $t_\alpha(n)$ 为 $t(n)$ 分布的上 α 分位点.

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, $t_\alpha(n)\approx z_\alpha$.



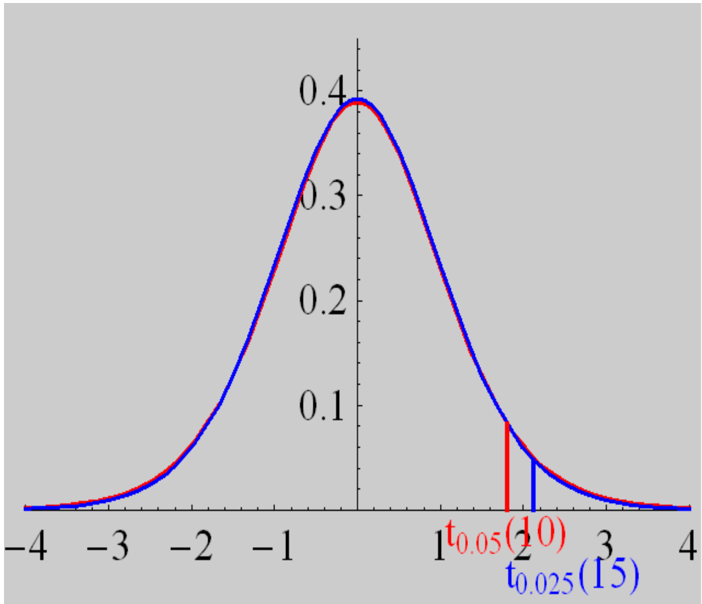
例3 设 $T \sim t(n)$, $t(n)$ 的上 α 分位点满足

$$P\{T > t_\alpha(n)\} = \int_{t_\alpha(n)}^{+\infty} t(y;n)\mathrm{d}y = \alpha,$$

求 $t_\alpha(n)$ 的值, 可通过查表完成.

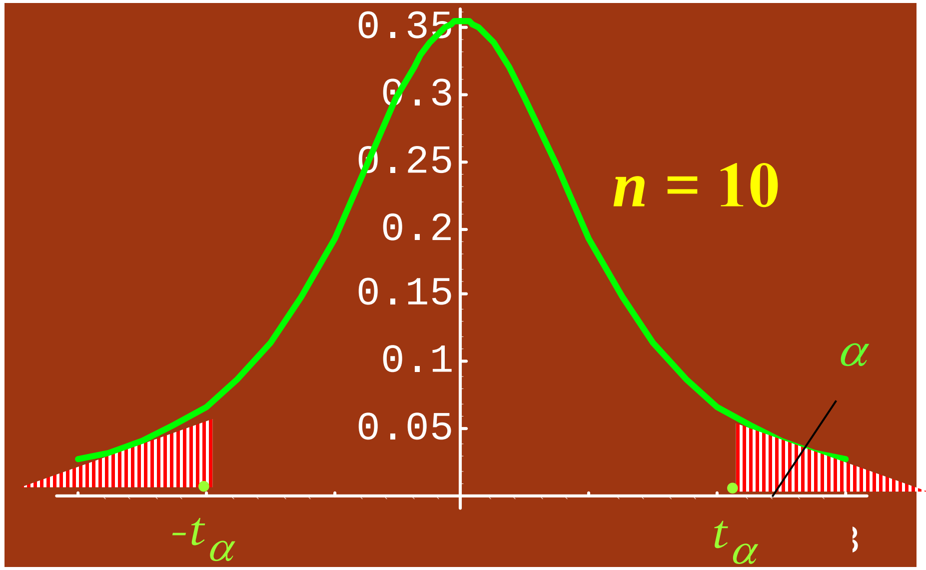
$t_{0.05}(10) = 1.8125,$

$t_{0.025}(15) = 2.1315.$



$$P(T > t_{\alpha}) = \alpha$$

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



$$P(T > 1.8125) = 0.05 \Rightarrow t_{0.05}(10) = 1.8125$$

$$P(T < -1.8125) = 0.05, \quad P(T > -1.8125) = 0.95$$

$$\Rightarrow t_{0.95}(10) = -1.8125$$

3. F 分布

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

n_1称为第自由度，n_2称为第二自由度。

$F(n_1, n_2)$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分布的数学期望为:

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

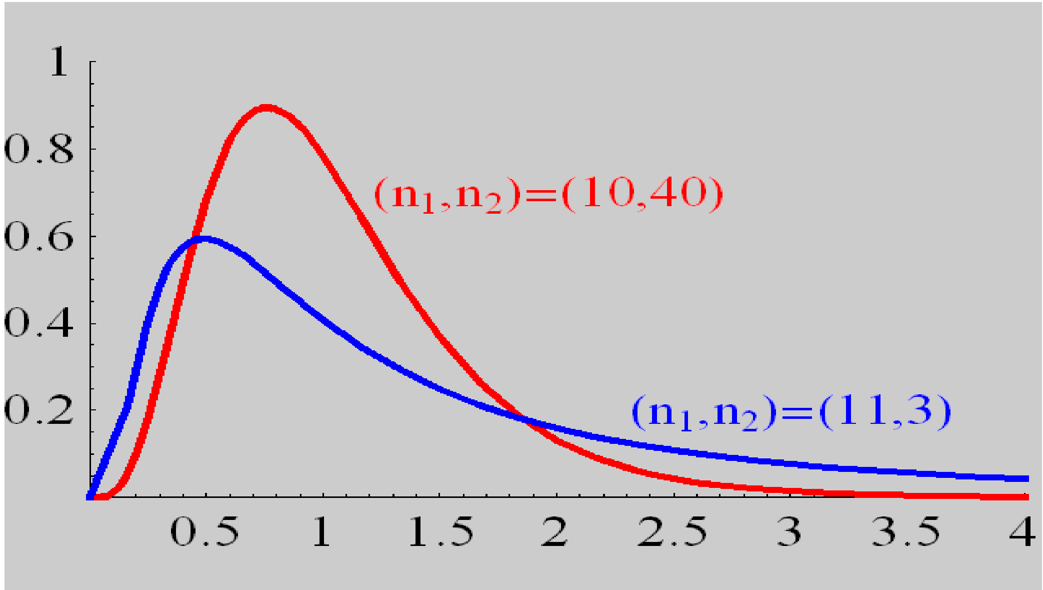
即它的数学期望并不依赖于第一自由度n_1.

F分布的概率密度曲线如图

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 分布的分位点

对于给定的α，$0 < \alpha < 1$，称满足条件

$$P\{F > F_\alpha(n_1, n_2)\} = \int_{F_\alpha(n_1, n_2)}^{+\infty} \psi(y)\mathrm{d}y = \alpha$$

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

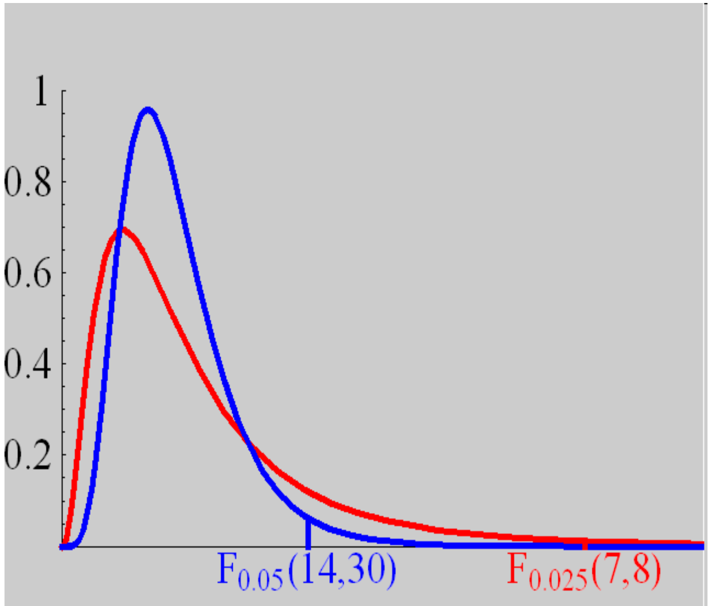
例4　设 $F(n_1, n_2)$分布的上α分位点满足

$$P\{F > F_\alpha(n_1, n_2)\} = \int_{F_\alpha(n_1, n_2)}^{+\infty} \psi(y)\mathrm{d}y = \alpha,$$

求 $F_\alpha(n_1, n_2)$的值,可通过查表完成.

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

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



F 分布的上 α 分位点具有如下性质：

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

证明　因为$F \sim F(n_1, n_2)$,

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

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

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

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

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

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

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

用来求分布表中未列出的一些上α分位点.

例 $F_{0.95}(12,9) = \dfrac{1}{F_{0.05}(9, 12)} = \dfrac{1}{0.28} = 0.357$.

4. 正态总体的样本均值与样本方差的分布

定理一 (样本均值的分布)

设 $X_1, X_2, \cdots, X_n$ 是来自正态总体 $N(\mu,\sigma^2)$ 的样本，$\bar{X}$ 是样本均值，S^2是样本方差 则有

$$E(\bar{X}) = \mu,$$

$$D(\bar{X}) = \sigma^2/n, \quad E(S^2) = \sigma^2$$

$$\bar{X} \sim N(\mu, \sigma^2/n).$$

即 $$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

注：

$$E(s^2)=E\left[\frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2-n\bar{X}^2\right)\right]$$

$$=\frac{1}{n-1}\left[\sum_{i=1}^{n}E(X_i^2)-nE(\bar{X}^2)\right]$$

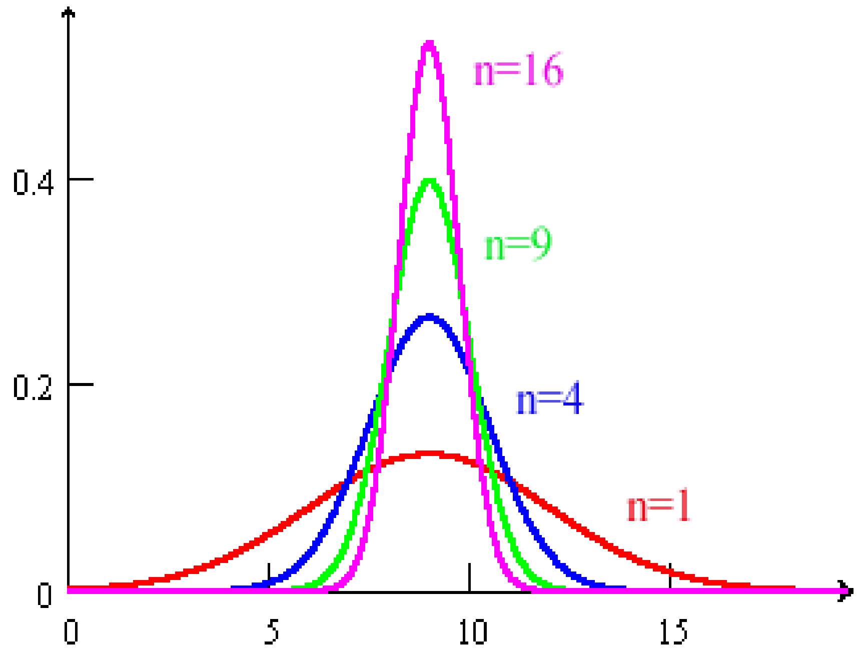
$$=\frac{1}{n-1}\left[\sum_{i=1}^{n}\left(\sigma^2+\mu^2\right)-n\left(\sigma^2/n+\mu^2\right)\right]$$

$$=\sigma^2$$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \Leftrightarrow \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

请注意 ：

n取不同值时样本均值 $\bar{X}$ 的分布



定理二

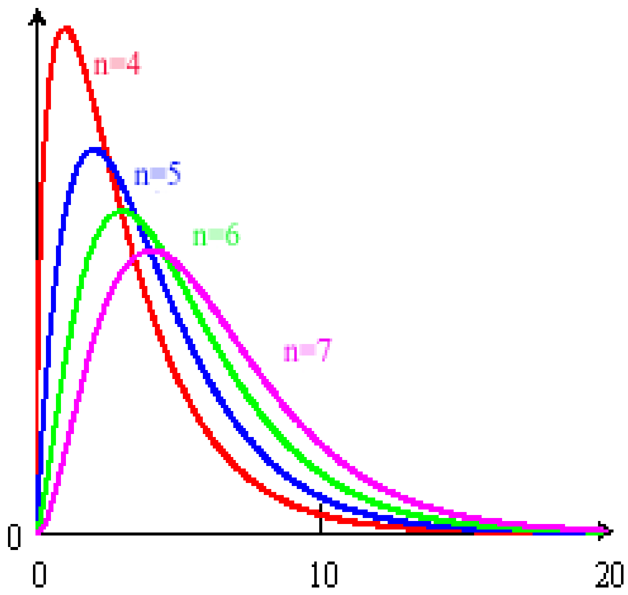
设 $X_1, X_2, \cdots, X_n$ 是总体 $N(\mu,\sigma^2)$ 的样本，

$\overline{X}$，S^2 分别是样本均值和样本方差，则有

(1) $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$;

(2) $\overline{X}$ 与 S^2 独立.

n取不同值时 $\dfrac{(n-1)S^2}{\sigma^2}$ 的分布



定理的证明见本章末附录.

定理三

设 $X_1, X_2, \cdots, X_n$ 是总体 $N(\mu,\sigma^2)$ 的样本,

$\overline{X}, S^2$ 分别是样本均值和样本方差, 则有

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1).$$

证明　因为 $\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$, $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$,

且两者独立, 由 t 分布的定义知

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \Bigg/ \sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}} \sim t(n-1).$$

定理四 (两总体样本均值差、样本方差比的分布)

设 $X_1, X_2, \cdots, X_{n_1}$ 与 $Y_1, Y_2, \cdots, Y_{n_2}$ 分别是具有相同方差的两正态总体 $N(\mu_1, \sigma_1^2)$, $N(\mu_2, \sigma_2^2)$ 的样本, 且这两个样本互相独立， 设 $\overline{X} = \dfrac{1}{n_1}\sum_{i=1}^{n_1} X_i$, $\overline{Y} = \dfrac{1}{n_2}\sum_{i=1}^{n_2} Y_i$ 分别是这两个样本的均值,

$$S_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(X_i - \overline{X})^2, \quad S_2^2 = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(Y_i - \overline{Y})^2$$

分别是这两个样本的方差, 则有

(1) $\dfrac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$;

(2) 当$\sigma_1^2=\sigma_2^2=\sigma^2$时,

$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \sim t(n_1+n_2-2),$$

其中 $S_w^2=\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$, $S_w=\sqrt{S_w^2}$.

证明 (1) 由定理二

$$\frac{(n_1-1)S_1^2}{\sigma_1^2} \sim \chi^2(n_1-1), \quad \frac{(n_2-1)S_2^2}{\sigma_2^2} \sim \chi^2(n_2-1),$$

由假设 S_1^2, S_2^2 独立, 则由 F 分布的定义知

$$\frac{(n_1-1)S_1^2}{(n_1-1)\sigma_1^2} \Bigg/ \frac{(n_2-1)S_2^2}{(n_2-1)\sigma_2^2} \sim F(n_1-1, n_2-1),$$

即 $$\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1-1, n_2-1).$$

(2) 因为 $\bar{X}-\bar{Y} \sim N\left(\mu_1-\mu_2, \frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}\right)$

所以 $U=\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim N(0,1)$,

由 $\frac{(n_1-1)S_1^2}{\sigma^2} \sim \chi^2(n_1-1)$, $\frac{(n_2-1)S_2^2}{\sigma^2} \sim \chi^2(n_2-1)$,

且它们相互独立，故由 χ^2 分布的可加性知

$$V=\frac{(n_1-1)S_1^2}{\sigma^2}+\frac{(n_2-1)S_2^2}{\sigma^2}\sim\chi^2(n_1+n_2-2),$$

由于 U 与 V 相互独立，按 t 分布的定义．

$$\frac{U}{\sqrt{V/(n_1+n_2-2)}}$$

$$=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2).$$

例5 设总体X服从正态分布$N(12,\sigma^2)$,抽取容量为25的样本,求样本均值$\bar{X}$大于12.5的概率.如果(1)已知$\sigma=2$; （2）σ未知，但已知样本方差$S^2=59.017$.

解 (1) $$P\{\bar{X}>12.5\}=P\left\{\frac{\bar{X}-12}{2/\sqrt{25}}>\frac{12.5-12}{2/\sqrt{25}}\right\}$$
$$=P\left\{\frac{\bar{X}-12}{0.4}>1.25\right\}=1-\Phi(1.25)=0.1063$$

(2) $$P\{\bar{X}>12.5\}=P\left\{\frac{\bar{X}-12}{S/\sqrt{25}}>\frac{12.5}{S/\sqrt{25}}\right\}=P\{T>1.059\}$$

查自由度为24的t分布表,$t_{0.15}(24)=1.059$,即

$P\{T>1.059\}=0.15$. 故有$P\{\bar{X}>12.5\}=0.15$.

例6　从正态总体$N(\mu,0.5^2)$中抽取样本$X_1,\cdots,X_{10}$.

（1）已知$\mu=0$，求概率$p\left\{\sum_{i=1}^{10}X_i^2\geq 4\right\}$;

（2）μ未知，求概率$p\left\{\sum_{i=1}^{10}(X_i-\bar{X})^2\geq 2.85\right\}$.

解　(1) 由$\mu=0$,有$X_i/0.5\sim N(0,1)$，则

$$Y^2\stackrel{\Delta}{=}\frac{1}{0.5^2}\sum_{i=1}^{10}X_i^2\sim\chi^2(10)$$

$$p\left\{\sum_{i=1}^{10}X_i^2\geq 4\right\}=p\left\{\frac{1}{0.5^2}\sum_{i=1}^{10}X_i^2\geq\frac{4}{0.5^2}\right\}=p\left\{Y^2\geq 16\right\}$$

查表求$\chi^2_{0.10}(10)=16$.由此可得 $p\left\{\sum_{i=1}^{10}X_i^2\geq 4\right\}=0.10.$

(2) 由题设及定理2,

$$Z\overset{\Delta}{=}\frac{9S^2}{0.5^2}=\frac{1}{0.5^2}\sum_{i=1}^{10}(X_i-\bar{X})^2\sim\chi^2(9)$$

$$p\left\{\sum_{i=1}^{10}(X_i-\bar{X})^2\geq 2.85\right\}=p\left\{\frac{1}{0.5^2}\sum_{i=1}^{10}(X_i-\bar{X})^2\geq\frac{2.85}{0.5^2}\right\}$$

$$=P\left\{Z^2\geq 11.4\right\}$$

查表得$\chi^2_{0.25}(9)=11.4$,由此可求得 **附表5查不到此值**

$$p\left\{\sum_{i=1}^{10}(X_i-\bar{X})^2\geq 2.85\right\}=0.25.$$

例7 设总体服从泊松分布$\pi(\lambda)$，$X_1,\cdots,X_n$是一个样本：

（1）写出$X_1,\cdots,X_n$的概率分布；

（2）计算$E(\bar{X}),D(\bar{X})$和$E(S^2)$.

解（1）由于$P\{X_i=x_i\}=\dfrac{\lambda^{x_i}}{x_i!}e^{-\lambda}\qquad x_i=0,1,2,\cdots,\lambda>0$

因此样本$X_1,\cdots,X_n$的概率分布为

$$\prod_{i=1}^{n}P\{X_i=x_i\}=\prod_{i=1}^{n}\frac{\lambda^{x_i}}{x_i!}e^{-\lambda}=e^{-n\lambda}\lambda^{\sum_{i=1}^{n}x_i}\Big/\prod_{i=1}^{n}x_i!$$

(2) 由于，$E(X)=D(X)=\lambda$,则有$E(\bar{X})=E(X)=\lambda$,

$$D(\bar{X})=\frac{D(X)}{n}=\frac{\lambda}{n}\quad E(S^2)=E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2\right]=\lambda$$

例8 若总体$X \sim N(0,1)$，从此总体中取一个容量为6的样本$X_1, X_2 \cdots, X_6$,设

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

试决定常数C，使随机变量CY服从χ^2分布.

解　因为 $X_1+X_2+X_3 \sim N(0,3)$

所以 $\dfrac{X_1+X_2+X_3}{\sqrt{3}} \sim N(0,1)$ 从而 $\left(\dfrac{X_1+X_2+X_3}{\sqrt{3}}\right)^2 \sim \chi^2(1)$

同理可知 $\left(\dfrac{X_4+X_5+X_6}{\sqrt{3}}\right)^2 \sim \chi^2(1)$

由χ^2分布的性质可知

$$\frac{1}{3}Y=\left(\frac{X_1+X_2+X_3}{\sqrt{3}}\right)^2+\left(\frac{X_1+X_2+X_3}{\sqrt{3}}\right)^2 \sim \chi^2(2) \quad \text{故} \ C=\frac{1}{3}.$$

小结

在这一节中我们学习了统计量的概念，几个重要的统计量及其分布，即抽样分布. 要求大家熟练地掌握它们.

常用的统计量

样本平均值 $$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

样本方差 $$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

样本标准差 $$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

样本k阶原点矩 $$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

样本k阶中心矩 $$B_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$$

抽样分布

χ^2分布 设$X_1,\cdots,X_n$相互独立,且均服从正态分布$N(0,1)$,

则称随机变量 $\chi^2=\sum_{i=1}^{n} X_i^2$服从自由度为$n$的$\chi^2$分布,

记为$\chi^2 \sim \chi^2(n)$.

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

F分布 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$F=\dfrac{U/n_1}{V/n_2}$服从自由度为（n_1,n_2）的分布,

记为$F \sim F(n_1,n_2)$.

抽样分布定理

样本均值的分布

设$X \sim N(\mu, \sigma^2)$，$X_1, \cdots, X_n$是来自总体X的样本，

则样本均值$\bar{X}$有$\bar{X} \sim N(\mu, \sigma^2/n)$.

样本方差、均值的分布

设$X_1, \cdots, x_n$是来自总体$N(\mu, \sigma^2)$的样本，$\bar{X}, S^2$

分别是样本均值和样本方差，则有

(1) $$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

(2) $\bar{X}$与S^2独立.

(3) $$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$$

两总体样本均值差、样本方差比的分布

设$X_1,\cdots,X_{n_1}$与$Y_1,\cdots,Y_{n_2}$分别来自总体$N(\mu_1,\sigma_1^2)$和$N(\mu_2,\sigma_2^2)$的样本，且这两个样本相互独立.$\bar{X},\bar{Y}$分别是这两个样本的样本均值；S_1^2,S_2^2分别是这两个样本的样本方差，则有

1、 $$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$$

2、 $$\frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}\sim t(n_1+n_2-2)$$

作业：课后习题 6、8、9

练习：

1、设总体X服从两点分布$b(1, p)$，其中p是未知参数，$X_1,\cdots,X_5$是来自X的简单随机样本.试指出X_1+X_2, $\max\limits_{1\le i\le 5} X_i, X_5+2p, (X_5-X_1)^2$之中哪些是统计量,哪些不是统计量，为什么？

解　$X_1+X_2, \max\limits_{1\le i\le 5} X_i, (X_5-X_1)^2$都是统计量，但

X_5+2p不是统计量（因为p是未知数）.

2、从有一个白球，两个黑球的罐子里有放回的取球，令$X=0$表示取到白球，$X=1$表示取到黑球，求容量为5的样本$X_1,\cdots,X_5$的和分布.并求样本的均值$\overline{X}$和样本的方差S^2的期望值.

解　X_i相互独立且都服从（0−1）分布$b(1,\dfrac{2}{3})$

令$Y=X_1+X_2+X_3+X_4+X_5$,则$Y\sim b(5,\dfrac{2}{3})$

其分布律为$P\{Y=k\}=C_5^k\left(\dfrac{2}{3}\right)^k\left(\dfrac{1}{3}\right)^{5-k},k=1,2,3,4,5$

由$E(X_i)=\dfrac{2}{3},D(X_i)=\dfrac{2}{9},i=1,2,3,4,5$

有　$E(\overline{X})=\dfrac{2}{3},E(S^2)=\dfrac{2}{9}.$

3、已知X ~ t(n)，那么X^2 ~______

A $F(1, n)$ B $F(n, 1)$ C $\chi^2(n)$ D $t(n)$

解 $\boldsymbol{X \sim t(n)}$，则 $X = \dfrac{U}{\sqrt{V/n}}$

其中 $\boldsymbol{U \sim N(0,1)}$，$\boldsymbol{V \sim \chi^2(n)}$

故 $\boldsymbol{X^2 = \dfrac{U^2/1}{V/n}} \sim F(1,n)$

4、总体$X \sim N(\mu,\sigma^2)$，$X_1,\cdots,X_n,X_{n+1}$为样本，

$\boldsymbol{\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i}$ $\boldsymbol{S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$ 求$\dfrac{X_{n+1}-\overline{X}}{S}\sqrt{\dfrac{n}{n+1}}$的分布

解 由$\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$，又$\overline{X} \sim N(\mu,\dfrac{\sigma^2}{n})$，

$X_{n+1} - \overline{X} \sim N\left(0,\dfrac{(n+1)\sigma^2}{n}\right)$ 故 $\boldsymbol{\dfrac{X_{n+1}-\overline{X}}{\sigma}\sqrt{\dfrac{n}{n+1}} \sim N(0,1)}$

于是 $\left(\dfrac{X_{n+1}-\overline{X}}{\sigma}\sqrt{\dfrac{n}{n+1}}\right)\bigg/\sqrt{\dfrac{(n-1)s^2}{(n-1)\sigma^2}}\sim t(n-1)$

即 $\dfrac{X_{n+1}-\overline{X}}{S}\sqrt{\dfrac{n}{n+1}}\sim t(n-1)$

5、证明：$[t(n)]^2=F(1,n)$

证 设 $X\sim t(n)$，$X=\dfrac{G}{\sqrt{\dfrac{\chi^2(n)}{n}}}$，$G\sim N(0,1)$

令

$$Y=X^2=\frac{G^2}{\dfrac{\chi^2(n)}{n}}=\frac{\dfrac{\chi^2(1)}{1}}{\dfrac{\chi^2(n)}{n}}\sim F(1,n)$$