四 概率分布的分位数

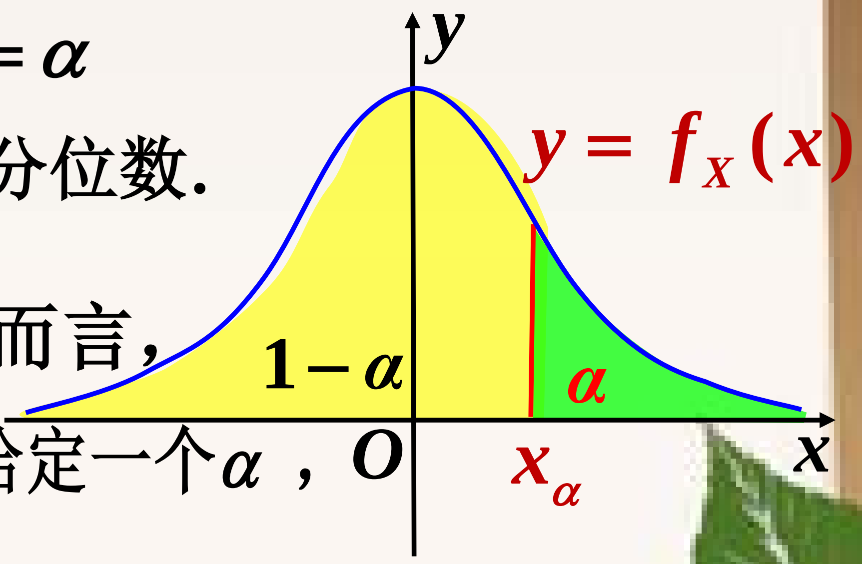
1. 定义

定义1.11 对于总体X和给定的$\alpha(0<\alpha<1)$,

若存在x_α,使

$$P\{X > x_\alpha\} = \alpha$$

则称x_α为X的分布的上侧α分位数.



注：(1) 对于连续型随机变量而言，α与x_α是一对一的。也即，给定一个α，存在唯一的x_α与之对应。

(2) 连续情形，x_α随着α的减小在增大，即x_α是α的单调减函数。

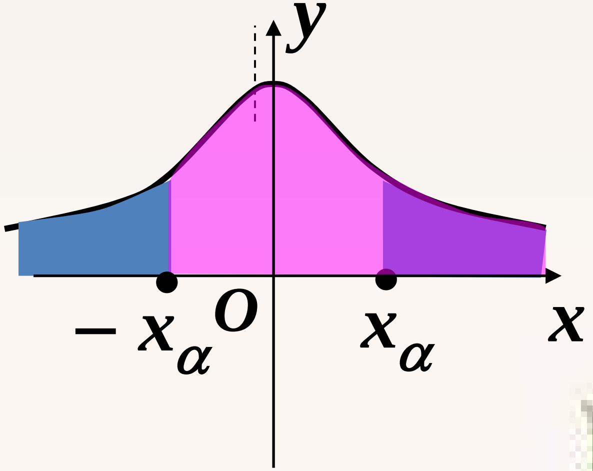
2. 常用分布的上侧分位数记号

	对称分布		非对称分布	
分布	$N(0,1)$	$t(n)$	$\chi^2(n)$	$F(n_1,n_2)$
记号	u_α	$t_\alpha(n)$	$\chi^2_\alpha(n)$	$F_\alpha(n_1,n_2)$

3. 查表法

(1) 若X的分布密度关于y轴对称，则

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



特例:

1) $N(0,1)$: $u_{1-\alpha} = -u_{\alpha}$

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

设 X 服从标准正态分布 $N(0,1)$， 则其上侧分位数 u_α 满足

$$P\{X > u_\alpha\} = \frac{1}{\sqrt{2\pi}}\int_{u_\alpha}^{+\infty} e^{-\frac{x^2}{2}}\,\mathrm{d}x$$

$$= 1 - P\{X \le u_\alpha\}$$

$$= 1 - \Phi(u_\alpha) = \alpha$$

即 $\Phi(u_\alpha) = 1 - \alpha$

给定α,由附表1可查得u_α的值.

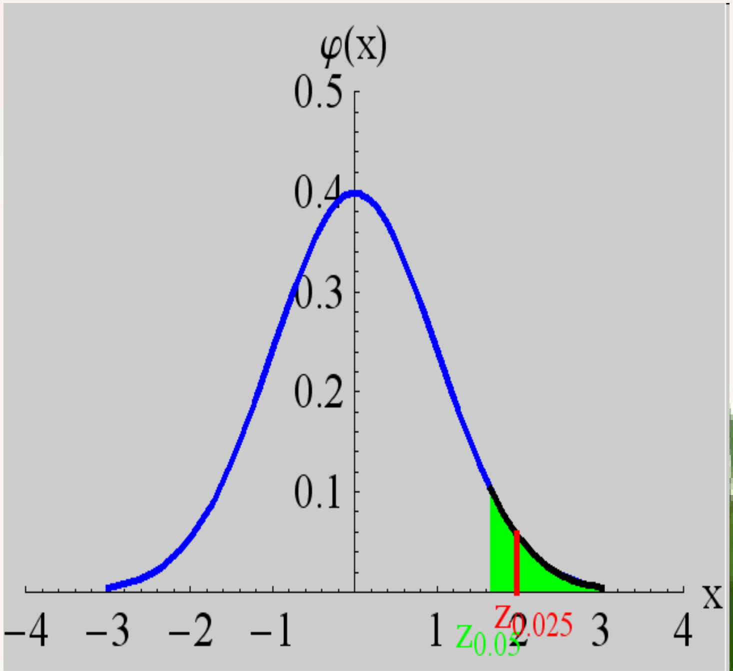
$$\boxed{\Phi(u_\alpha) = 1 - \alpha}$$

$u_{0.05} = 1.645,$ 附表1 **0.95** $(\alpha = 0.05)$

$u_{0.025} = 1.96,$ 附表1 **0.975** $(\alpha = 0.025)$

根据正态分布的对称性知

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



2) t分布的上侧分位数$t_\alpha(n)$:

对于给定的α, $0<\alpha<1$, 称满足条件

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

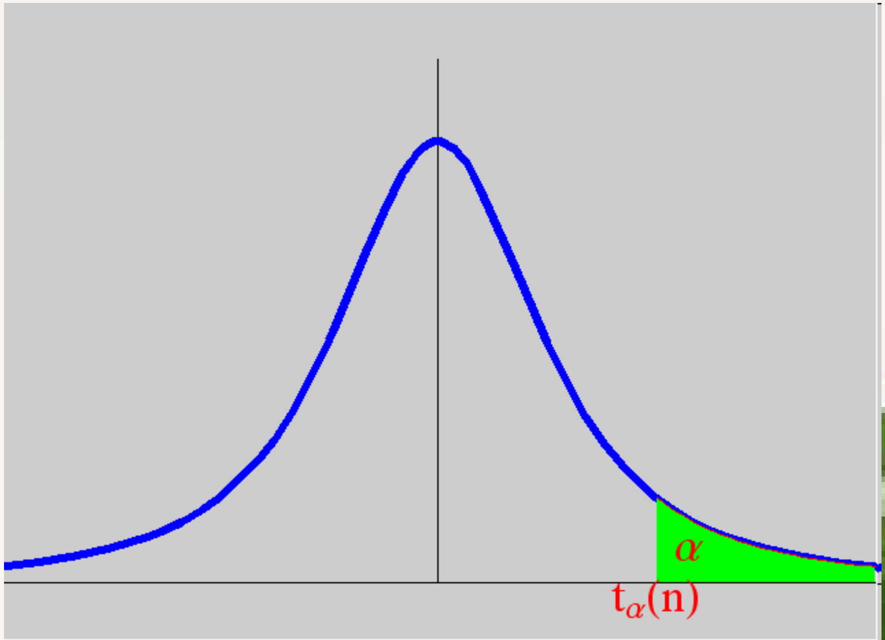
的点$t_\alpha(n)$为$t(n)$分布的上α分位点.

可以通过查表求得上α分位点的值.

由分布的对称性知

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

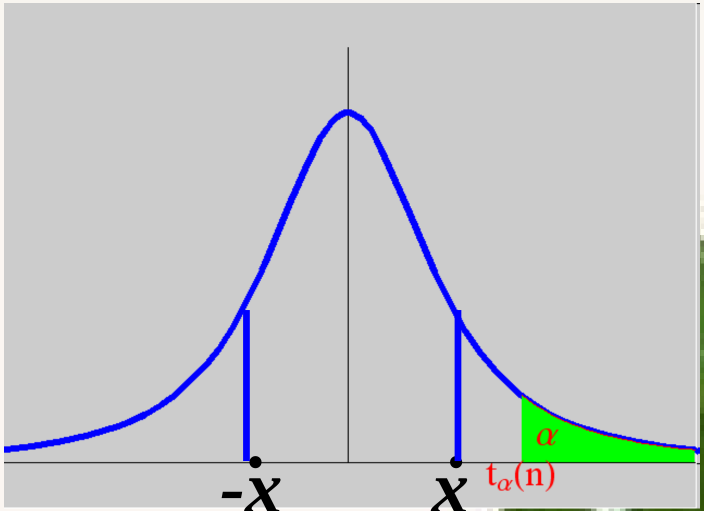
当$n>45$时, $t_\alpha(n)\approx u_\alpha$.



例1 设随机变量T服从t分布$t(n)$, $n>1$, 设给定$\alpha(0<\alpha<1)$时，实数$t_\alpha(n)$满足$P\{T>t_\alpha(n)\}=\alpha$,则当$P\{|T|<x\}=\alpha$时，$x=$ __$t_{(1-\alpha)/2}(n)$

解 $P\{|T|<x\}=\alpha \Rightarrow P\{|T|>x\}=1-\alpha$

$\Rightarrow P\{T>x\}=(1-\alpha)/2$



(2) 若X的分布密度无对称性，

1) $\chi^2_\alpha(n)$:对于给定的正数α，$0<\alpha<1$，称满足

$$P\{\chi^2>\chi^2_\alpha(n)\}=\int_{\chi^2_\alpha(n)}^{\infty}p(y)dy=\alpha$$

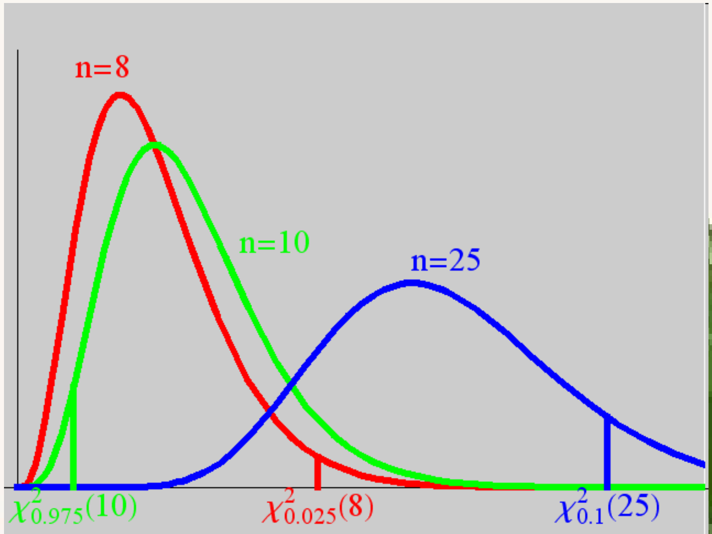
的点$\chi^2_\alpha(n)$为$\chi^2(n)$分布的上侧分位数.

当$n\le 60$时，可查表 3　(表3只详列到 n=60 为止).

$\chi^2_{0.025}(8)=17.535$,　附表3

$\chi^2_{0.975}(10)=3.247$,　附表3

$\chi^2_{0.1}(25)=34.382$.　附表3



当$n>60$时，$\chi_{\alpha}^{2}(n)\approx n+\sqrt{2n}\,u_{\alpha}$.

费歇(R.A.Fisher)公式：

费歇资料

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

其中u_{α}是标准正态分布的上α分位点.

例如：$\chi_{0.05}^{2}(120)\approx 120+\sqrt{2\times 120}\times u_{0.05}$

$$=120+\sqrt{240}\times 1.64$$

$$=145.5.$$

2) $F_\alpha(n_1,n_2)$: 对于 $\alpha=0.01,\ 0.025,\ 0.05,\ 0.1$ 等，

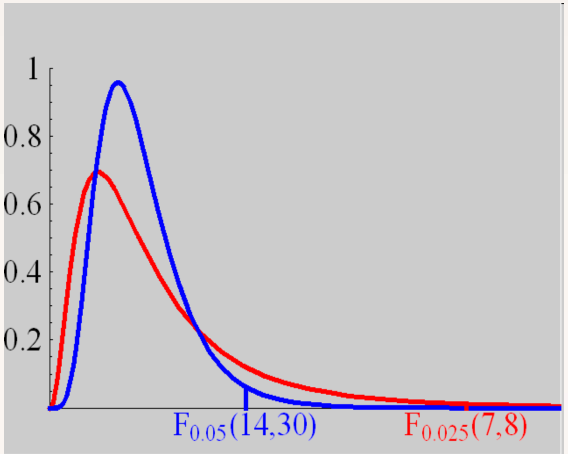
可直接查表4.1 ~ 4.4.

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

附表 4.1

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

附表 4.2



此外，还可利用关系

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

由F_α求得$F_{1-\alpha}$.

如： $$F_{0.95}(12,9)=\frac{1}{F_{0.05}(9,12)}=\frac{1}{2.8}=0.357.$$

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

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

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

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

例2 设随机变量X服从F分布$F(8,8)$，若$\alpha>0$为已知常数，且满足条件$P(X>\alpha)=0.10$,则$P(X>1/\alpha)=$___0.9___.

解 $X\sim F(8,8)$, $\Rightarrow$ $1/X\sim F(8,8)$, $\Rightarrow$ X与$1/X$同分布。

$P(X>\alpha)=0.10$ $\Rightarrow$ $P(1/X<1/\alpha)=0.10$,

$\Rightarrow$ $P(1/X>1/\alpha)=0.9$, $\Rightarrow$ $P(X>1/\alpha)=0.9$

五、正态总体样本均值和方差的分布

1. 单个总体样本均值的分布

定理1.11

设 $X_1, X_2, \cdots, X_n$ 来自正态总体$N(\mu, \sigma^2)$的样本,即$X_i \sim N(\mu,\sigma^2)\quad (i=1,2,\cdots,n)$

则它们的任一确定的线性函数

$$U=\sum_{i=1}^{n} a_i X_i \sim N(\mu\sum_{i=1}^{n}a_i, \sigma^2\sum_{i=1}^{n}a_i^2).$$

其中$a_1,a_2,\cdots,a_n$为不全为零的常数.

特别当$a_i = \frac{1}{n}(i = 1, 2, \cdots, n)$时，可以得到$\bar{X}$的分布

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

2. 单个总体样本方差的分布

定理1.12*** 设 $X_1, X_2, \cdots, X_n$ 是总体 $N(\mu, \sigma^2)$ 的样本, $\bar{X}, S_n^2$ 分别是样本均值和样本方差, 则有

(1) $\dfrac{nS_n^2}{\sigma^2} = \dfrac{(n-1)S_n^{*2}}{\sigma^2} \sim \chi^2(n-1)$;

(2) $\bar{X}$ 与 S_n^2(或S_n^{*2}) 独立.

证

令$Y_i = \dfrac{X_i - \mu}{\sigma}$,则$Y_i \sim N(0,1)$且$Y_1, Y_2, \cdots Y_n$相互独立,

故

$$\boldsymbol{Y} = (Y_1, Y_2, \cdots, Y_n)^T \sim N(\boldsymbol{0}, \mathbf{I}_n)$$

构造正交矩阵T使得,

$$T = \begin{pmatrix} \dfrac{1}{\sqrt{n}} & \dfrac{1}{\sqrt{n}} & \cdots & \dfrac{1}{\sqrt{n}} \\ t_{21} & t_{22} & \cdots & t_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ t_{n1} & t_{n2} & \cdots & t_{nn} \end{pmatrix}$$

设$Z=(Z_1,Z_2,\cdots,Z_n)^T=TY$,则$E(Z)=0$,

$$\mathrm{cov}(Z,Z)=\mathrm{cov}(TY,TY)$$

$$=T\,\mathrm{cov}(Y,Y)T^T=TT^T=I_n$$

所以$Z\sim N(0,\mathrm{I}_n)$

则$$Z_1=(\frac{1}{\sqrt{n}},\frac{1}{\sqrt{n}},\cdots,\frac{1}{\sqrt{n}})Y=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Y_i=\sqrt{n}\bar{Y}=\sqrt{n}\,\frac{\bar{X}-\mu}{\sigma}$$

$$\frac{nS_n^2}{\sigma^2}=\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2=\sum_{i=1}^{n}(\frac{X_i-\bar{X}}{\sigma})^2$$

$$=\sum_{i=1}^{n}(\frac{X_i-\mu+\mu-\bar{X}}{\sigma})^2=\sum_{i=1}^{n}(\frac{X_i-\mu}{\sigma})^2-n(\frac{\bar{X}-\mu}{\sigma})^2$$

$$\frac{nS_n^2}{\sigma^2}=\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2-n\left(\frac{\bar{X}-\mu}{\sigma}\right)^2$$

$$=\sum_{i=1}^{n}Y_i^2-n\bar{Y}^2=\sum_{i=1}^{n}Y_i^2-(\sqrt{n}\bar{Y})^2$$

$$=Z^TTT^TZ-(Z_1)^2=Z^TZ-(Z_1)^2$$

$$=\sum_{i=1}^{n}Z_i^2-(Z_1)^2=\underline{\sum_{i=2}^{n}}Z_i^2\sim\chi^2(n-1)$$

不含Z_1，和Z_1独立。

又由于

$$\bar{X}=\sigma\bar{Y}+\mu=\frac{\sigma}{\sqrt{n}}Z_1+\mu$$

由此可以看到，$\bar{X}$与$\frac{nS_n^2}{\sigma^2}$相互独立.

定理1.13 设$(X_1, X_2, \cdots, X_n)$是总体$N(\mu,\sigma^2)$的样本，$\bar{X}, S_n^{*2}$分别是样本均值和修正样本方差,则有

$$T = \frac{\bar{X}-\mu}{S_n^{*} / \sqrt{n}} = \frac{\bar{X}-\mu}{S_n / \sqrt{n-1}} \sim t(n-1).$$

证 $\because U = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1), V = \dfrac{(n-1)S_n^{*2}}{\sigma^2} \sim \chi^2(n-1),$

且两者独立, 由 t 分布的定义知

$$T = \frac{U}{\sqrt{\frac{V}{n-1}}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \Bigg/ \sqrt{\frac{(n-1)S_n^{*2}}{\sigma^2(n-1)}} \sim t(n-1).$$

3. 两个正态总体样本均值差的分布

定理1.14 设总体 $X \sim N(\mu_1, \sigma_1^2)$， $Y \sim N(\mu_2, \sigma_2^2)$，$X$与$Y$相互独立. 样本$(X_1, X_2, \cdots, X_{n_1})$ 与 $(Y_1, Y_2, \cdots, Y_{n_2})$ 分别来自 总体X和Y，则

(1) $\bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2})$

或 $$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2 / n_1 + \sigma_2^2 / n_2}} \sim N(0,1)$$

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

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

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

$S_{1n_1}^{*^2}$和$S_{2n_2}^{*^2}$分别是来自两个总体样本的修正样本方差.

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

下证（2）

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

由 $\frac{(n_1-1)S_{1n_1}^{*2}}{\sigma^2} \sim \chi^2(n_1-1),$

$$\frac{(n_2-1)S_{2n_2}^{*2}}{\sigma^2} \sim \chi^2(n_2-1),$$

且它们相互独立, 故由 χ^2 分布的可加性知

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

由于 U 与 V 相互独立,按 t 分布的定义

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

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

4. 两个正态总体样本方差商的分布

定理1.15

设总体 $X \sim N(\mu_1, \sigma_1^2)$， $Y \sim N(\mu_2, \sigma_2^2)$，

X与Y相互独立. 样本$(X_1, X_2, \cdots, X_{n_1})$

与$(Y_1, Y_2, \cdots, Y_{n_2})$分别来自总体$X$和$Y$，则

$$F = \frac{S_{1n_1}^{*2} / \sigma_1^2}{S_{2n_2}^{*2} / \sigma_2^2} \sim F(n_1 - 1, n_2 - 1);$$

证 $\dfrac{(n_1-1)S_{1n_1}^{*2}}{\sigma_1^2} \sim \chi^2(n_1-1),\ \dfrac{(n_2-1)S_{2n_2}^{*2}}{\sigma_2^2} \sim \chi^2(n_2-1),$

由假设 $S_{1n_1}^{*2}, S_{2n_2}^{*2}$ 独立，则由 F 分布的定义知

$$\frac{(n_1-1)S_{1n_1}^{*2}}{(n_1-1)\sigma_1^2} \Bigg/ \frac{(n_2-1)S^{*2}_{2n_2}}{(n_2-1)\sigma_2^2} \sim F(n_1-1, n_2-1),$$

即 $F = \dfrac{S_{1n_1}^{*2}/\sigma_1^2}{S_{2n_2}^{*2}/\sigma_2^2} \sim F(n_1-1, n_2-1).$

六、一些非正态总体样本均值的分布

1. 问题的提出

抽样分布的精确分布可以归属到计算随机变量或随机向量函数的分布，但是从关于随机变量或随机向量函数的分布介绍中可以看到计算相当复杂，因而对于一般总体情形下的抽样分布的计算几乎无法完成，因而对于一般情形，我们一方面可以考虑特殊总体情形下的精确抽样分布，另一方面考虑大样本情形下抽样分布的渐近分布。

2. 特殊情形下抽样分布 的精确分布

例4(p26例1.14) 设总体$X \sim B(m, p), X_1, X_2, \cdots, X_n$来自总体的一个样本，试求$\bar{X}$的分布.

证 由二项分布的可加性可知

$$n\bar{X} = \sum_{i=1}^{n} X_i \sim B(nm, p)$$

因此

$$P\{\bar{X} = \frac{k}{n}\} = P\{n\bar{X} = k\} = C_{nm}^{k} p^k (1-p)^{nm-k}$$

例5(p26例1.15) 设总体$X \sim P(\lambda), X_1, X_2, \cdots, X_n$来自总体的一个样本，试求$\overline{X}$的分布.

证 由泊松分布的可加性可知

$$n\overline{X} = \sum_{i=1}^{n} X_i \sim P(n\lambda)$$

因此

$$P\{\overline{X} = \frac{k}{n}\} = P\{n\overline{X} = k\} = \frac{(n\lambda)^k}{k!} \mathrm{e}^{-n\lambda}, k = 0,1,2,\cdots$$

例6(p27例1.16) 设总体X服从参数为λ的指数分布$\exp(\lambda)$, $X_1, X_2, \cdots, X_n$来自总体的一个样本，试求$\bar{X}$的分布.

证 由于指数分布是$\Gamma(1,\lambda)$，因而由其可加性可知

$$Y = n\bar{X} = \sum_{i=1}^{n} X_i \sim \Gamma(n,\lambda)$$

因此 $$f_Y(y) = \frac{\lambda^n}{\Gamma(n)} y^{n-1} e^{-\lambda y}, \quad y > 0,$$

故 $$f_{\bar{X}}(x) = f_T(nx)\,|(nx)'|$$

$$= \frac{(n\lambda)^n}{\Gamma(n)} x^{n-1} e^{-n\lambda x}, \quad x > 0,$$

Thank You!