河海大学

Ch6 样本及抽样分布

数理统计实际上是概率论的具体应用。它的研究范围分成两个方面，一个是统计推断，另一个是抽样理论与试验设计。本课程仅研究第一个方面的内容。统计推断主要研究抽样分布、参数估计、假设检验等。

- **基本概念**

1. 总体与样本

总体 研究对象的全体。

通常指研究对象的某项数量指标，

个体 组成总体的单元。

通常也指与总体对应的某项数量指标

样本 来自总体的部分个体。

n称为样本容量

总体 $X \sim f(x)$

样本 $X_1, \cdots, X_n$ n称为样本容量

又称其是“简单随机样本”或简称为“随机样本”或“样本”。

满足以下两个条件：

（1）独立性： $X_1, \cdots, X_n$ 相互独立；

（2）同分布性： $X_1, \cdots, X_n$ 与总体 X 同分布。

来自总体 X 的随机样本 $X_1, \cdots, X_n$ 可记为

$$X_1, \cdots, X_n \overset{iid}{\sim} X \text{ 或 } X_1, \cdots, X_n \overset{iid}{\sim} f(x)$$

其中 $f(x)$ 是 X 的概率函数。

样本观测值 对样本 $X_1, \cdots, X_n$进行观测，即可得一组观测值 $x_1, \cdots, x_n$。

- **统计量**

样本 $X_1, \cdots, X_n$的函数 $g(X_1, \cdots, X_n)$称为是总体 X 的一个**统计量**，若$g(X_1, \cdots, X_n)$与任何未知参数无关。

统计量的观测值

若样本 $X_1, \cdots, X_n$的观测值为$x_1, \cdots, x_n$，则$g(x_1, \cdots, x_n)$称为统计量$g(X_1, \cdots, X_n)$的观测值。

- 常用统计量

1. 样本均值（样本平均数）$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$

其观测值为 $\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$

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

其观测值为 $s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2$

样本均方差（标准差） $S=\sqrt{S^2}$

其观测值为 $s=\sqrt{s^2}$

3. k 阶样本矩

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

观测值为 $a_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$

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

观测值为 $b_k = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^2$

4. 极大、极小统计量

极大统计量 $X_{(n)}=\max\{X_1, \cdots, X_n\}$,

其观测值 $x_{(n)}=\max\{x_1, \cdots, x_n\}$;

极小统计量 $X_{(1)}=\min\{X_1, \cdots, X_n\}$,

其观测值 $x_{(1)}=\min\{x_1, \cdots, x_n\}$。

- **抽样分布**

统计量的分布称为**抽样分布**。

数理统计中主要研究如下四个分布：

U—分布、 χ^2—分布、 t —分布和F—分布。

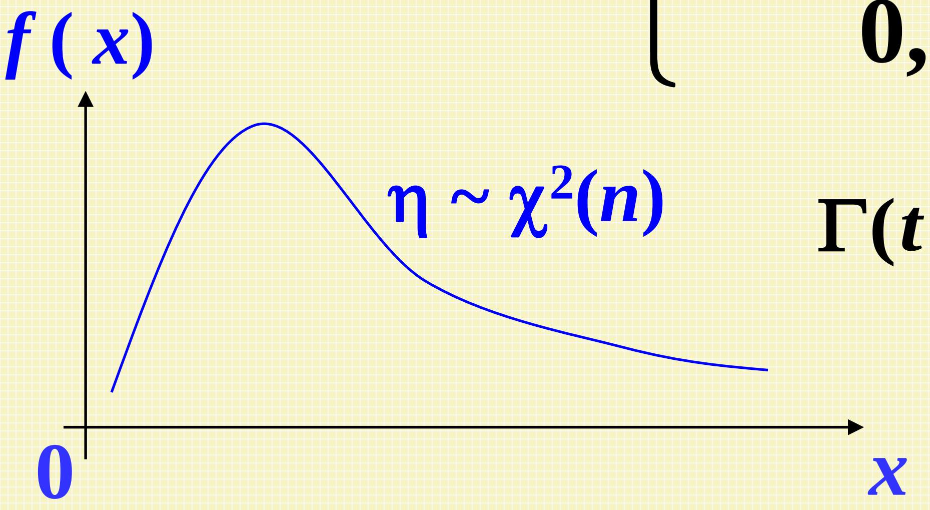
1. χ^2—分布

构造 设 $X_1,\cdots,X_n \overset{iid}{\sim} N(0,1)$, 则

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

$\chi^2(n)$称为自由度为 n 的χ^2-分布.

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



$$\Gamma(t)=\int_0^{+\infty}x^{t-1}e^{-x}dx,\ t>0$$

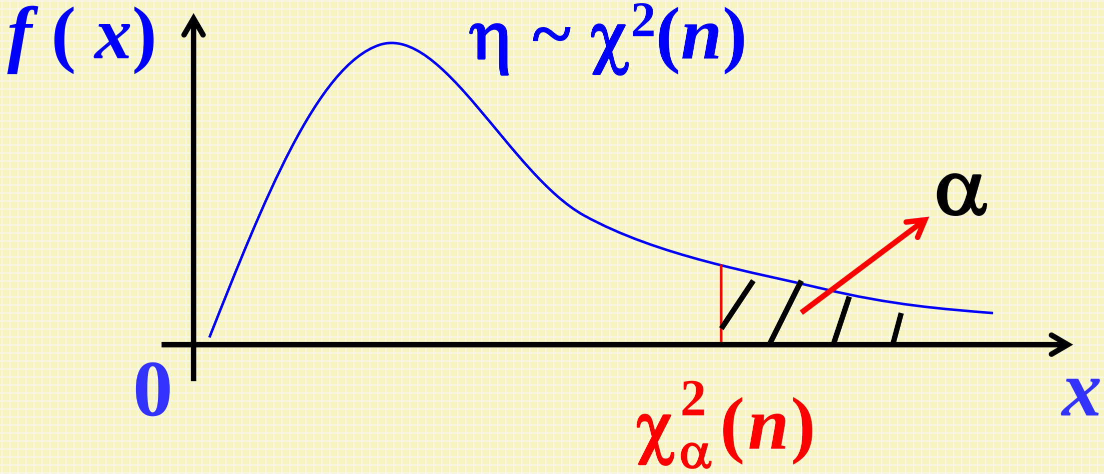
$$\Gamma(n)=n!$$

再生性 若$\eta_1 \sim \chi^2(n_1)$, $\eta_2 \sim \chi^2(n_2)$, η_1 , η_2独立，

则 $\eta_1 + \eta_2 \sim \chi^2(n_1+n_2)$。

期望与方差 若$\eta \sim \chi^2(n)$，则$E(\eta)= n$，$D(\eta)=2n$。

分位点 设$\eta \sim \chi^2(n)$，若对于α：$0< \alpha <1$， 存在$\chi_\alpha^2(n) > 0$， 满足$P\{\eta \geq \chi_\alpha^2(n)\} = \alpha$,则称$\chi_\alpha^2(n)$为$\chi^2(n)$的上侧$\alpha$分位点;



费歇尔(R.A.Fisher)曾经证明：当n充分大时，近似地有

$$\sqrt{2\eta} \sim N(\sqrt{2n-1}, 1).$$

其中$\eta \sim \chi^2(n)$，从而当n充分大时(一般地 > 45)，近似地有

$$\chi_\alpha^2(n) = \frac{1}{2}(z_\alpha + \sqrt{2n-1})^2.$$

其中z_α为$N(0，1)$的上侧α分位点。

2. t—分布

构造 若 $\xi \sim N(0, 1)$, $\eta \sim \chi^2(n)$, ξ与η独立，则

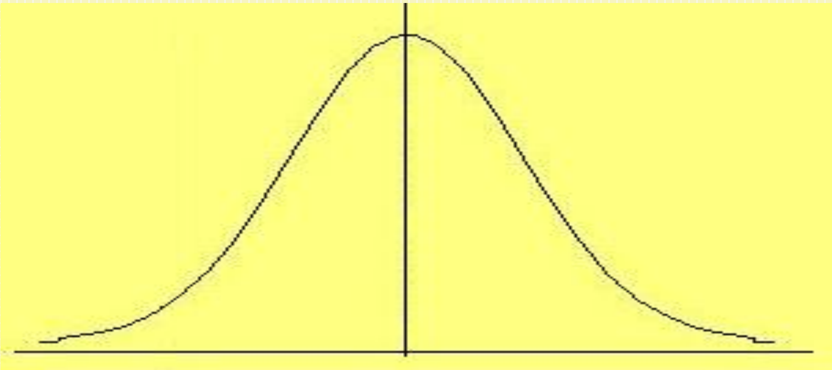
$$T = \frac{\xi}{\sqrt{\eta / n}} \sim t(n).$$

$t(n)$称为自由度为n的 t—分布

其密度为

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

密度函数 $f(t)$的图形与 $N(0, 1)$的密度函数的图形很象，只是 $t(n)$的图形两端尾巴厚一些。



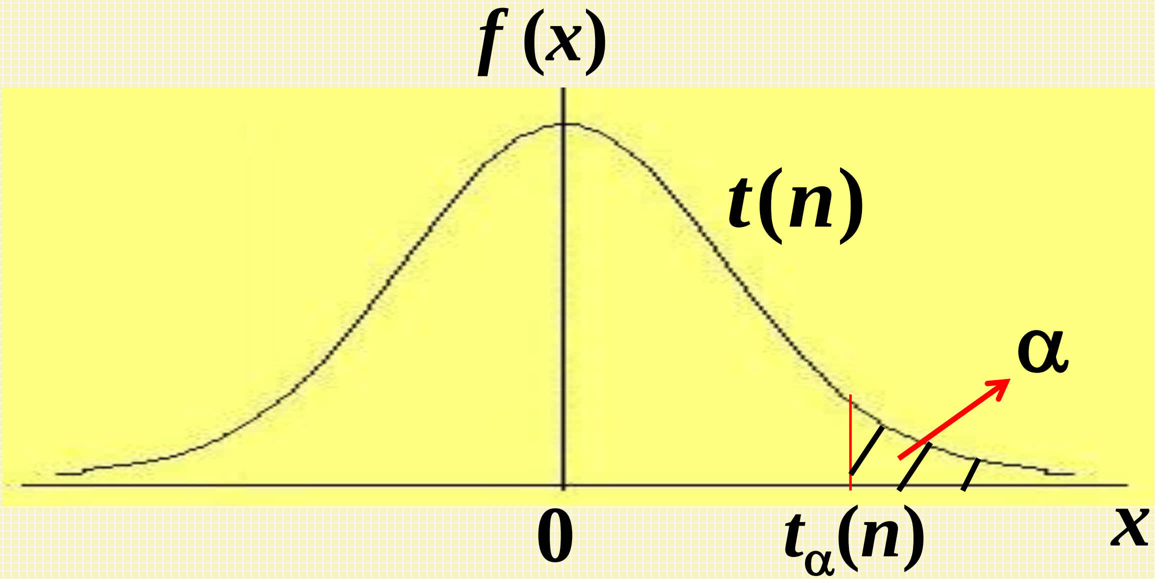
基本性质

(1) $f(t)$关于$t=0$(纵轴)对称。

(2) $f(t)$的极限为$N(0，1)$的密度函数，即

$$\lim_{n\to\infty} f(t)=\varphi(t)=\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}},-\infty<x<\infty$$

分位点 设$T \sim t(n)$，若对α：$0<\alpha<1$，存在$t_\alpha(n)>0$，满足$P\{T \geq t_\alpha(n)\}=\alpha$，则称$t_\alpha(n)$为$t(n)$的**上侧分位点**



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

- F—分布

构造 若$\eta_1 \sim\chi^2(n_1)$，$\eta_2\sim\chi^2(n_2)$，η_1，η_2独立，则

$$F=\frac{\eta_1/n_1}{\eta_2/n_2}\sim F(n_1,n_2).$$

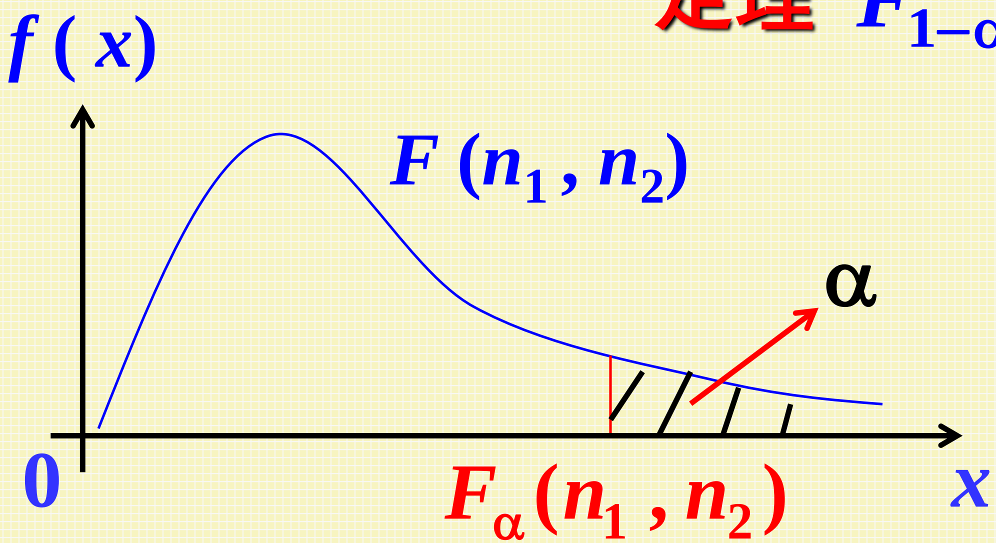
$F(n_1, n_2)$称为第一自由度为n_1，第二自由度为n_2的F—分布。

定理 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α：0<α<1，若存在 $F_\alpha(n_1, n_2)> 0$，满足$P\{F \geq F_\alpha(n_1, n_2)\}=\alpha$，则称 $F_\alpha(n_1, n_2)$为$F(n_1, n_2)$的上侧α分位点。

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



• 正态总体的抽样分布

定理 若$X_1,\cdots,X_n$为取自$N(\mu,\sigma^2)$的容量为n的样本，

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i,\quad S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2.$$ 则有

(1) $\overline{X}\sim N(\mu,\frac{\sigma^2}{n})$, 即 $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$;

(2) $\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$; (3) $\overline{X}$与S^2相互独立;

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

定理 若 $X_1, X_2, \cdots, X_{n_1} \overset{iid}{\sim} N(\mu_1, \sigma_1^{\ 2})$，

$Y_1, Y_2, \cdots, Y_{n_2} \overset{iid}{\sim} N(\mu_2, \sigma_2^{\ 2})$，且两样本独立

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

$$\overline{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i, \quad S_2^{\ 2} = \frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(Y_i - \overline{Y})^2;$$ 则有

$$(1)\ \overline{X}-\overline{Y} \sim N\left(\mu_1-\mu_2, \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)$$

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

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

(3)进一步地，假定$\sigma_1^2=\sigma_2^2$，就有，

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

其中

$$S_w^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$称为混合样本方差.

例 设总体$X \sim N(\mu, \sigma^2)$,如果要以99.7%的概率保证偏差$|\overline{X}-\mu|<0.1$,试问在$\sigma^2=0.5$时,样本容量n应取多大?

例 设$X_1,\cdots,X_9$是来自正态总体X的简单随机样本,令$Y_1=\frac{1}{6}\sum_{i=1}^{6}X_i$, $Y_2=\frac{1}{3}(X_7+X_8+X_9)$,

$$S_2=\frac{1}{2}\sum_{i=7}^{9}(X_i-Y_2)^2,\quad Z=C\frac{Y_1-Y_2}{S}$$

问题C取什么值时Z服从t分布,自由度为多少?

例　设$(X_1,\cdots,X_n)$是来自正态总体 $N(0,\sigma^2)$的样本$(\sigma>0)$,求:

$$(1)\ P\{\frac{\sigma^2}{2}<\frac{1}{16}\sum_{i=1}^{16}(X_i-\mu)^2\le 2\sigma^2\};$$

$$(2)\ P\{\frac{\sigma^2}{2}<\frac{1}{16}\sum_{i=1}^{16}(X_i-\overline{X})^2\le 2\sigma^2\}.$$