▲概率论的基本概念

珩玉

①加法公式（Jordan公式）

- 对于任意事件A.B ：$P(A\cup B)=P(A)+P(B)-P(AB)$
- 对于任意n个事件$A_1,A_2\cdots A_n$：$P(A_1\cup A_2\cup\cdots\cup A_n)=\sum_{i=1}^{n}P(A_i)-\sum_{1\le i<j\le n}P(A_iA_j)+\sum_{1\le i<j<k\le n}P(A_iA_jA_k)+\cdots+(-1)^{n-1}P(A_1A_2\cdots A_n)$ [奇加偶减]

[例] $P(A_1\cup A_2\cup A_3)=P(A_1)+P(A_2)+P(A_3)-P(A_1A_2)-P(A_1A_3)-P(A_2A_3)+P(A_1A_2A_3)$

②古典概型

- 不放回的有顺序的排列 $A_n^k=n!/(n-k)!$
- 不放回的无顺序的组合 $C_n^k=n!/(n-k)!k!$ $[C_n^{r_1}\cdot C_{n-r_1}^{r_2}\cdot C_{n-r_1-r_2}^{r_3}\cdots C_{r_k}^{r_k}=n!/r_1!\,r_2!\cdots r_k!]$

③条件概率

设A.B为2个事件，且$P(A)>0$，称 $P(B|A)=P(AB)/P(A)$ 为条件概率

④乘法定理（乘法公式）

- $P(AB)=P(B|A)\cdot P(A)$ $(P(A)>0)$
- $P(ABC)=P(C|AB)\cdot P(B|A)\cdot P(A)$ $(P(A)\geqslant P(AB)>0)$
- $P(A_1A_2\cdots A_n)=P(A_n|A_1A_2\cdots A_{n-1})P(A_{n-1}|A_1A_2\cdots A_{n-2})\cdots P(A_2|A_1)P(A_1)$

⑤全概率公式

设试验E的样本空间为S，A为E的事件，$B_1,B_2,\cdots B_n$为S的一个划分 且$P(B_i)>0$ $(i=1,2,\cdots n)$则，

$P(A)=P(A|B_1)P(B_1)+P(A|B_2)P(B_2)+\cdots+P(A|B_n)P(B_n)$

⑥贝叶斯公式

$$P(B_i|A)=P(B_iA)/P(A)=\frac{P(A|B_i)\,P(B_i)}{\sum_{j=1}^{n}P(A|B_j)\,P(B_j)}\quad ,i=1,2,\cdots n$$

⑦事件的独立性

[定义]：若$P(AB)=P(A)P(B)$，则称A.B两事件相互独立

- n个事件相互独立，则对其中任意2个.3个…n个事件的积事件的概率，等于各事件概率之积
- $A_1,A_2,\cdots A_n$相互独立，则$A_1,A_2\cdots A_n$中至少有一个发生的概率 $P(\bigcup_{i=1}^{n}A_i)=1-P(\overline{\bigcup_{i=1}^{n}A_i})=1-P(\bigcap_{i=1}^{n}\overline{A_i})=1-\prod_{i=1}^{n}P(\overline{A_i})$

▲随机变量及其分布

①分布函数

$F(x)=P\{X\le x\}$

[性质]
- $F(x)$为一个不减函数，$F(x_2)-F(x_1)=P\{x_1<X\le x_2\}\geqslant0$
- $0\le F(x)\le1$
- $F(x)$右连续

②概率密度函数 [典型例题 P_{43} 例1]（P_{51} 例2.例3）

$F(x)=\int_{-\infty}^{x}f(t)\,dt$ [$f(x)$称为X的概率密度函数]

[性质]
- $\int_{-\infty}^{+\infty}f(x)\,dx=1$
- $P\{x_1<X\le x_2\}=F(x_2)-F(x_1)=\int_{x_1}^{x_2}f(x)\,dx$
- 若$f(x)$在x点连续，则$F'(x)=f(x)$
- $f(x)\geqslant0$

③求解函数分布函数 [典型例题 P_{53} 例5]

设随机变量X具有概率密度$f_X(x)$，又设函数$g(x)$处处可导且恒有$g'(x)>0$(或恒有$g'(x)<0$)

则，$f_Y(y)=\begin{cases}f_X[h(y)]\,|h'(y)| & ,\alpha<y<\beta\\ 0 & ,其他\end{cases}$

其中，$\alpha=\min\{g(-\infty),g(+\infty)\}$，$\beta=\max\{g(-\infty),g(+\infty)\}$

$h(y)$为$g(x)$的反函数

④分布及其概率

■离散型变量分布

琐玉

二项分布 $X\sim b(n,p)$ $P(X=k)=p^k(1-p)^{n-k}C_n^k$

(参数为λ)泊松分布 $X\sim \pi(\lambda)$ $P(X=k)=\frac{\lambda^k e^{-\lambda}}{k!}$

几何分布 $P(X=k)=(1-p)^{k-1}p$

超几何分布 $X\sim H(n,N,M)$ $P(X=m)=C_M^m C_{N-M}^{n-m}/C_N^n$

[$m=0,1\cdots l$，$l=\min(M,n)$]

[典型例题 P_{38} 例5]

●泊松定理（二项分布的泊松逼近）[用于n很大，p很小时]

设λ>0是一个常数，n是任意正整数，设 $np_n=\lambda$，则对于任一固定的非负整数k，有

$\lim\limits_{n\to\infty} C_n^k p_n^k(1-p_n)^{n-k}=\lambda^k e^{-\lambda}/k!$

■连续型随机变量分布

均匀分布 $X\sim U(a,b)$ $f(x)=\begin{cases}\frac{1}{b-a}, & a<x<b\\ 0, & \text{其他}\end{cases}$ $F(x)=\begin{cases}0, & x<a\\ \frac{x-a}{b-a}, & a\le x<b\\ 1, & x\ge b\end{cases}$

指数分布 $X\sim$ 参数为θ的指数分布 $f(x)=\begin{cases}\frac{1}{\theta}e^{-x/\theta}, & x>0\\ 0, & \text{其他}\end{cases}$ $F(x)=\begin{cases}1-e^{-x/\theta}, & x>0\\ 0, & \text{其他}\end{cases}$ 其中，θ>0为常数

正态分布 $X\sim N(\mu,\sigma^2)$ $f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty<x<+\infty$ $F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dt, -\infty<x<+\infty$

标准正态分布函数 $\phi(x)$

- $\phi(-x)=1-\phi(x)$
- $P(|x|\le a)=2\phi(a)-1$
- $P(X\le a)=\phi\left(\frac{a-\mu}{\sigma}\right)$
- $P(x>a)=1-\phi\left(\frac{a-\mu}{\sigma}\right)$

▲多维随机变量及其分布

①二维随机变量

分布函数 $F(x,y)=P\{(X\le x)\cap(Y\le y)\}\overset{\text{记为}}{=}P\{X\le x, Y\le y\}$ [则，$P(x_1<X<x_2, y_1<Y<y_2)=F(x_2,y_2)+F(x_1,y_1)-F(x_1,y_2)-F(x_2,y_1)$]

概率密度 $f(x,y)$

- $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)\,dx\,dy=F(\infty,\infty)=1$
- 若 $f(x,y)$ 在 (x,y) 连续，则 $\frac{\partial^2 F(x,y)}{\partial x\partial y}=f(x,y)$
- 设G为xoy面上区域，点 (X,Y) 落在G内的概率为 $P\{(X,Y)\in G\}=\iint_G f(x,y)\,dx\,dy$

②边缘分布 [典例 P_{66} 例2]

$F_X(x)=F(x,\infty)$；$F_Y(y)=F(\infty,y)$ 边缘分布函数

$p_{i\cdot}=\sum_{j=1}^{\infty}p_{ij}=P(X=x_i), i=1,2\cdots$；$p_{\cdot j}=\sum_{i=1}^{\infty}p_{ij}=P(Y=y_j), j=1,2\cdots$ 边缘分布律

$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\,dy$；$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)\,dx$ 边缘概率密度

③条件分布 [典例 P_{71} 例3 例4]

$P\{X=x_i|Y=y_j\}=P\{X=x_i,Y=y_j\}/P(Y=y_j)$

$P\{Y=y_j|X=x_i\}=P\{X=x_i,Y=y_j\}/P(X=x_i)$

●二维均匀分布 $f(x,y)=\begin{cases}\frac{1}{A}, & (x,y)\in G\\ 0, & \text{其他}\end{cases}$ （其中，A为G的面积）

④相互独立

[定义] 若 $F(x,y)=F_X(x)F_Y(y)$，则称 x,y 相互独立 $\overset{\text{等价}}{\Longleftrightarrow}$ $f(x,y)=f_X(x)f_Y(y)$ 在平面上几乎处处成立（除去面积为0的集合）

([推论] 二维正态分布随机变量 (X,Y)，X和Y相互独立 $\overset{\text{等价}}{\Longleftrightarrow}$ 关联参数 $\rho=0$)

[定理] 若 $(X_1,X_2,\cdots X_m)$ 和 $(Y_1,Y_2,\cdots Y_n)$ 相互独立，则 $X_i\ (i=1,2\cdots)$ 和 $Y_j\ (j=1,2\cdots)$ 相互独立。

若h,g为连续函数，则，$h(X_1,X_2\cdots X_m)$ 和 $g(Y_1,Y_2\cdots Y_n)$ 相互独立。

⑤ 2个随机变量的函数分布　　　　　　　　　　　　　　　　　　　　衍玉

■ $Z=X+Y$ 的分布 [典例 P_{87} 24]

一般 $\begin{cases} f_{X+Y}(z)=\int_{-\infty}^{+\infty} f(z-y,y)\,dy \\ f_{X+Y}(z)=\int_{-\infty}^{+\infty} f(x,z-x)\,dx \end{cases}$　　X,Y相互独立 $\begin{cases} f_{X+Y}(z)=\int_{-\infty}^{+\infty} f_X(z-y)f_Y(y)\,dy \\ f_{X+Y}(z)=\int_{-\infty}^{+\infty} f_X(x)f_Y(z-x)\,dx \end{cases}$

[推论] 有限个相互独立的正态随机变量的线性组合仍服从正态分布

即 $C_1X_1+C_2X_2+\cdots+C_nX_n \sim N\left(\sum_{i=1}^{n} C_i\mu_i,\ \sum_{i=1}^{n} C_i^2\sigma_i^2\right)$

■ $Z=\frac{Y}{X}$ 的分布

一般 $f_{Y/X}(z)=\int_{-\infty}^{+\infty} |x| f(x,xz)\,dx$　　X,Y相互独立 $f_{Y/X}(z)=\int_{-\infty}^{+\infty} |x| f_X(x) f_Y(xz)\,dx$

■ $Z=XY$ 的分布

一般 $f_{XY}(z)=\int_{-\infty}^{+\infty} \frac{1}{|x|} f(x,\frac{z}{x})\,dx$　　X,Y相互独立 $f_{XY}(z)=\int_{-\infty}^{+\infty} \frac{1}{|x|} f_X(x) f_Y(\frac{z}{x})\,dx$

■ $M=\max\{X,Y\}$ 的分布 [典例 P_{88} 29]

X,Y相互独立 $F_{max}(z)=F_X(z)F_Y(z)$　　n维，相互独立 $F_{max}(z)=F_{X_1}(z)F_{X_2}(z)\cdots F_{X_n}(z)$　　n维，独立同分布 $F_{max}(z)=[F(z)]^n$

■ $N=\min\{X,Y\}$ 的分布

X,Y相互独立 $F_{min}(z)=1-[1-F_X(z)]\cdot[1-F_Y(z)]$　　n维，相互独立 $F_{min}(z)=1-[1-F_{X_1}(z)]\cdot[1-F_{X_2}(z)]\cdots[1-F_{X_n}(z)]$　　n维，独立同分布 $F_{min}(z)=1-[1-F(z)]^n$

• 随机变量分布（2种特殊变量）

ⓐ 若X,Y相互独立，$X\sim\pi(\lambda_1)$，$Y\sim\pi(\lambda_2)$ 则 $Z=X+Y\sim\pi(\lambda_1+\lambda_2)$

ⓑ 若X,Y相互独立，$X\sim b(n_1,p)$，$Y\sim b(n_2,p)$ 则 $Z=X+Y\sim b(n_1+n_2,p)$

▲随机变量的数字特征

①数学期望 [典例 P_{114} 7]

离散型 $E(X)=\sum_{k=1}^{\infty} x_k p_k$　　$E(Y)=E[g(X)]=\sum_{k=1}^{\infty} g(x_k)p_k$　　$E(Z)=E[g(X,Y)]=\sum_{j=1}^{\infty}\sum_{i=1}^{\infty} g(x_i,y_j)p_{ij}$

连续型 $E(X)=\int_{-\infty}^{+\infty} x f(x)\,dx$　　$E(Y)=E[g(X)]=\int_{-\infty}^{+\infty} g(x)f(x)\,dx$　　$E(Z)=E[g(X,Y)]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)\,dx\,dy$

~~二维随机变量 $E(Z)=E[g(X,Y)]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)\,dx\,dy$~~

[性质] • $E(CX)=CE(X)$　　• 对于任意X,Y，$E(X+Y)=E(X)+E(Y)$　　• 若X,Y相互独立 $E(XY)=E(X)E(Y)$

②方差 [典例 P_{116} 22]

[定义] $D(X)=E(X^2)-[E(X)]^2$　　离散型 $D(X)=\sum_{k=1}^{\infty}[x_k-E(X)]^2 p_k$　　连续型 $D(X)=\int_{-\infty}^{+\infty}[x-E(X)]^2 f(x)\,dx$

[性质] • $D(CX)=C^2D(X)$，$D(X+C)=D(X)$　　• 对于任意X,Y，$D(X\pm Y)=D(X)+D(Y)\pm 2E\{(X-E(X))(Y-E(Y))\}$　　• 若X,Y相互独立 $D(X\pm Y)=D(X)+D(Y)$

• $D(X)=0 \Leftrightarrow P\{X=E(X)\}=1$

★常见概率分布表

		参数		
(0-1)分布	$P\{X=k\}=p^k(1-p)^{1-k}$，$k=0,1$	$0<p<1$	$E(X)=p$	$D(X)=p(1-p)$
二项分布 $X\sim b(n,p)$	$P\{X=k\}=C_n^k p^k(1-p)^{n-k}$，$k=0,1,2\cdots n$	$\left(\begin{matrix} n\geq 1 \\ 0<p<1 \end{matrix}\right)$	$E(X)=np$	$D(X)=np(1-p)$
几何分布	$P\{X=k\}=(1-p)^{k-1}p$，$k=1,2\cdots$	$0<p<1$	$E(X)=\frac{1}{p}$	$D(X)=(1-p)/p^2$
泊松分布 $X\sim\pi(\lambda)$	$P\{X=k\}=\lambda^k e^{-\lambda}/k!$，$k=0,1,2\cdots$	$\lambda>0$	$E(X)=\lambda$	$D(X)=\lambda$
均匀分布 $X\sim U(a,b)$	$f(x)=\begin{cases}\frac{1}{b-a}, & a<x<b \\ 0, & 其他\end{cases}$	$a<b$	$E(X)=\frac{a+b}{2}$	$D(X)=\frac{(b-a)^2}{12}$
正态分布 $X\sim N(\mu,\sigma^2)$	$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/(2\sigma^2)}$	μ, $\sigma>0$	$E(X)=\mu$	$D(X)=\sigma^2$
指数分布	$f(x)=\begin{cases}\frac{1}{\theta}e^{-x/\theta}, & x>0 \\ 0, & 其他\end{cases}$	$\theta>0$	$E(X)=\theta$	$D(X)=\theta^2$

③切比雪夫不等式

设随机变量X有$E(X)=\mu$，$D(X)=\sigma^2$ 则对于任意正数ε，不等式 $P\{|X-\mu|\geq\varepsilon\}\leq\frac{\sigma^2}{\varepsilon^2}$ 成立 　　　珩玉

④协方差及相关系数 [典例 P_{117} 32]

协方差 $Cov(X,Y)=E\{[X-E(X)][Y-E(Y)]\}$
$=E(XY)-E(X)E(Y)$

[性质] • $Cov(X,Y)=Cov(Y,X)=E(XY)-E(X)E(Y)$
• $Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)$
• $Cov(X,X)=D(X)$
• $Cov(aX,bY)=ab\,Cov(X,Y)$
• $D(X\pm Y)=D(X)+D(Y)\pm2Cov(X,Y)$
• $Cov(a,X)=0$ （a为常数）

相关系数 $\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$ （$\rho_{XY}=0$时称X,Y不相关）

[性质] • $|\rho_{XY}|\leq1$ 　• $|\rho_{XY}|=1\Leftrightarrow P\{Y=a+bX\}=1$

[推论] 对于二维正态随机变量(X,Y)来说，X和Y相互独立 $\Leftrightarrow$ X和Y不相关

⑤矩，协方差矩阵

X的k阶原点矩（k阶矩） $\mu_k=E(X^k)$，$k=1,2\cdots$ 　一阶原点矩 $E(X)$

X的k阶中心矩 $\sigma_k=E\{[X-E(X)]^k\}$，$k=2,3\cdots$ 　二阶中心矩 $D(X)$

X和Y的k+l阶混合矩 $\mu_{kl}=E(X^kY^l)$，$k,l=1,2,\cdots$

X和Y的k+l阶混合中心矩 $\sigma_{kl}=E\{[X-E(X)]^k[Y-E(Y)]^l\}$，$k,l=1,2,\cdots$ 　二阶混合中心矩 $Cov(X,Y)$

协方差矩阵 $C=\begin{bmatrix}c_{11}&c_{12}&\cdots&c_{1n}\\c_{21}&c_{22}&\cdots&c_{2n}\\\vdots&\vdots&&\vdots\\c_{n1}&c_{n2}&\cdots&c_{nn}\end{bmatrix}$ （对称矩阵） 其中 $c_{ij}=Cov(X_i,X_j)=E\{[X_i-E(X_i)]\cdot[X_j-E(X_j)]\}$

▲大数定律及中心极限定理

①依概率收敛

[定义] 设$Y_1,Y_2,\cdots Y_n,\cdots$是一个随机变量序列，a为一个常数，若对任意正数$\varepsilon$有$\lim_{n\to\infty}P\{|Y_n-a|<\varepsilon\}=1$，则称序列$Y_1,Y_2,\cdots Y_n\cdots$依概率收敛于a，记为 $Y_n\xrightarrow{P}a$

[性质] 设 $X_n\xrightarrow{P}a$，$Y_n\xrightarrow{P}b$，$g(x,y)$在(a,b)处连续，则 $g(X_n,Y_n)\xrightarrow{P}g(a,b)$

②大数定理

弱大数定理：设随机变量$X_1,X_2,\cdots X_n,\cdots$相互独立，同分布，且$E(X_k)=\mu$，则序列$\bar{X}=\frac{1}{n}\sum_{k=1}^{n}X_k$依概率收敛于$\mu$，即 $\bar{X}\xrightarrow{P}\mu$

伯努利大数定理：设f_A是n次独立重复实验中事件A发生的次数，p是A发生的概率，则对任意正数$\varepsilon>0$，有$\lim_{n\to\infty}P\{|\frac{f_A}{n}-p|<\varepsilon\}=1$

③中心极限定理 [典例 P_{126} 3 　P_{127} 7]

[独立同分布的中心极限定理] 设随机变量$X_1,X_2,\cdots X_n\cdots$相互独立，同分布，且$E(X_k)=\mu$，$D(X_k)=\sigma^2>0$ $(k=1,2,\cdots)$

则，当n充分大时，$\frac{\sum_{k=1}^{n}X_k-n\mu}{\sqrt{n}\sigma}\overset{近似地}{\sim}N(0,1)$ 　[即 $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\overset{近似地}{\sim}N(0,1)$ 或 $\bar{X}\overset{近似地}{\sim}N(\mu,\sigma^2/n)$]

[二项分布的正态近似] 设随机变量η_n $(n=1,2,\cdots)$服从参数为n,p $(0<p<1)$的二项分布，则对任意x，有

$$\lim_{n\to\infty}P\{\frac{\eta_n-np}{\sqrt{np(1-p)}}\leq x\}=\Phi(x)$$

▲样本及抽样分布

①常见统计量

样本平均值 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ （观察值 $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$）

样本方差 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2=\frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2-n\overline{X}^2\right)$ （观察值 $s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2=\frac{1}{n-1}\left(\sum_{i=1}^{n}x_i^2-n\bar{x}^2\right)$）

样本标准差 $S=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}$ （观察值 $s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$） [算术平均值的A类不确定度 $u_A=\frac{S_x}{\sqrt{n}}=\sqrt{\frac{1}{n-1}\cdot\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i-\bar{x})^2}$]

样本k阶(原点)矩 $A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$, $k=1,2\cdots$ （观察值 $a_k=\frac{1}{n}\sum_{i=1}^{n}x_i^k$）

样本k阶中心矩 $B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k$, $k=2,3\cdots$ （观察值 $b_k=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^k$）

②三种典型分布 [典例 P_{167} 题4]

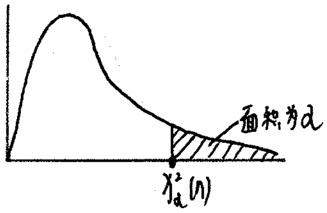
■ χ^2 分布

[定义] 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 $\chi^2\sim\chi^2(n)$

[性质] • 可加性：设 $\chi_1^2\sim\chi^2(n_1)$，$\chi_2^2\sim\chi^2(n_2)$ 且 χ_1^2 和 χ_2^2 相互独立，则有 $\chi_1^2+\chi_2^2\sim\chi^2(n_1+n_2)$

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

[分位点]



$\chi^2_\alpha(n)$ 为 $\chi^2(n)$ 分布的上 α 分位点

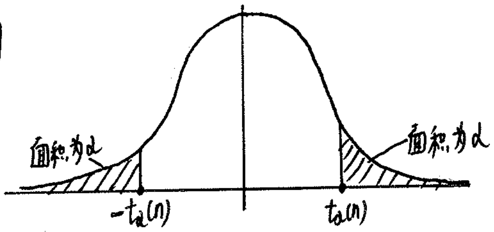
■ t 分布

[定义] 设 $X\sim N(0,1)$，$Y\sim\chi^2(n)$ 且 X, Y 相互独立，则称随机变量 $t=\frac{X}{\sqrt{Y/n}}$ 服从自由度为n的t分布，记为 $t\sim t(n)$

[性质] • 具有良好的对称性(概率密度)，$E(t)=0$

• 当n充分大时，图形类似于标准正态变量概率密度的图形

[分位点]



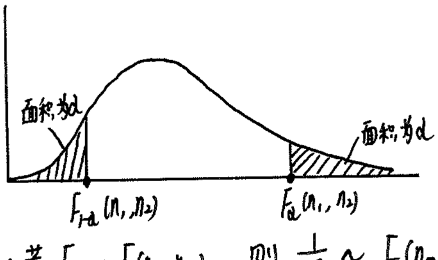
$t_\alpha(n)$ 为 $t(n)$ 分布的上 α 分位点，

$-t_\alpha(n)$ 亦称 $t(n)$ 分布的下 α 分位点

■ F 分布

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

[分位点]



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

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

[性质] • 若 $F\sim F(n_1,n_2)$ 则 $\frac{1}{F}\sim F(n_2,n_1)$

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

③几种重要统计量的分布　　　　　　　　　　　　　　　　　　　　　珩玉

- 设 $X_1, X_2, \cdots X_n$ 是来自正态总体 $N(\mu, \sigma^2)$ 的样本，$\bar{X}$ 为样本均值，S^2 为样本方差　(一个样本)

[定理一] $\mu \cdot \sigma^2$ 均已知，　$Z = \dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

[定理二] μ 已知，σ^2 未知，　$t = \dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$　[引入S变量,自由度为n-1]

[定理三] $\mu \cdot \sigma^2$ 均未知，　$\chi^2 = \dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$　；$\bar{X}$ 与 S^2 相互独立　；$E(S^2) = n D(\bar{X}) = \sigma^2$

- 设 $X_1, X_2, \cdots X_{n_1}$ 与 $Y_1, Y_2, \cdots, Y_{n_2}$ 分别是来自正态总体 $N(\mu_1, \sigma_1^2)$ 和 $N(\mu_2, \sigma_2^2)$，且这两个样本相互独立　(2个样本)

设 $\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$，$\bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$ 分别为样本均值；$S_1^2 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(X_i-\bar{X})^2$，$S_2^2 = \frac{1}{n_2-1}\sum_{i=1}^{n_2}(Y_i-\bar{Y})^2$ 为样本方差

[定理四] σ_1^2, σ_2^2 已知　$Z = \dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$

[定理五] $\sigma_1^2 \cdot \sigma_2^2$ 未知　$t = \dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_W\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim t(n_1+n_2-2)$　其中，$S_W = \dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$

[定理六] μ_1, μ_2 未知　$F = \dfrac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$

▲ 参数估计

①估计量　$\hat{\theta}(X_1, X_2, \cdots, X_n)$

估计值　$\hat{\theta}(x_1, x_2, \cdots, x_n)$

②矩估计法　[典例 P_{173} 2]

[例] $f(x) = \begin{cases} \theta c^{\theta} x^{-(\theta+1)}, & x>c \\ 0, & 其他 \end{cases}$　其中 $c>0$ 为已知，$\theta>1$，θ 为未知参数，求 θ 的矩估计量和矩估计值

解：$\mu_1 = \int_{-\infty}^{+\infty} x f(x)\,dx = \int_c^{+\infty} x\theta c^{\theta} x^{-(\theta+1)}dx = \theta c^{\theta}\int_c^{+\infty} x^{-\theta}dx = \theta c^{\theta}\dfrac{x^{1-\theta}}{1-\theta}\Big|_c^{+\infty} = \dfrac{\theta}{1-\theta}c^{\theta}\cdot c^{1-\theta} = \dfrac{\theta \cdot c}{1-\theta}$

解得，$\theta = \dfrac{\mu_1}{\mu_1 - c}$

以 $\bar{X}$ 代 μ_1 得，θ 的矩估计值为 $\hat{\theta} = \dfrac{\bar{x}}{\bar{x}-c}$，$\theta$ 的矩估计量为 $\hat{\theta} = \dfrac{\bar{X}}{\bar{X}-c}$

③最大似然估计法　[典例 P_{174} 7]

[例] 设 $X \sim b(1,p)$，$X_1, X_2, \cdots, X_n$ 是来自 X 的一个样本，试求参数 p 的最大似然估计量

解：设 $x_1, x_2 \cdots x_n$ 是相应于样本 $X_1, X_2, \cdots X_n$ 的一个样本值。

X 的分布律为 $P\{X=x\} = p^x(1-p)^{1-x}$，$x=0,1$

故似然函数为 $L(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$

$\ln L(p) = \sum_{i=1}^{n} x_i \ln p + (n - \sum_{i=1}^{n} x_i)\ln(1-p)$

令 $\dfrac{\ln L(p)}{dp} = \dfrac{1}{p}\cdot\sum_{i=1}^{n} x_i - \dfrac{1}{1-p}\left(n - \sum_{i=1}^{n} x_i\right) = 0$

解得，p 的最大似然估计值 $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$　　∴ p 的最大似然估计量 $\hat{p} = \bar{X}$

④估计量的评选标准 [典例 P_{175} 12]

- 无偏性　若估计量 $\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots X_n)$ 的数学期望 $E(\hat{\theta})$ 存在，且对任意 $\theta \in \Theta$ 有 $E(\hat{\theta}) = \theta$，则称 $\hat{\theta}$ 是 θ 的无偏估计量
- 有效性　设 $\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2 \cdots X_n)$ 与 $\hat{\theta}_2 = \hat{\theta}_2(X_1, X_2 \cdots X_n)$ 都是 θ 的无偏估计量，若对于任意的 $\theta \in \Theta$，有 $D(\hat{\theta}_1) < D(\hat{\theta}_2)$ 则称 $\hat{\theta}_1$ 较 $\hat{\theta}_2$ 有效
- 相合性　设 $\hat{\theta}(X_1, X_2 \cdots X_n)$ 为参数 θ 的估计量，若对于任意 $\theta \in \Theta$，当 $n\to+\infty$ 时，$\hat{\theta}(X_1, X_2 \cdots X_n)$ 依概率收敛于 θ，则称 $\hat{\theta}$ 为 θ 的相合估计量

[即 对于任意 $\varepsilon>0$，有 $\lim_{n\to+\infty} P\{|\hat{\theta}-\theta|<\varepsilon\} = 1$，则称 $\hat{\theta}$ 是 θ 的相合估计量]

⑤区间估计 [典例P35 16]　　　　班玉

• 置信水平

若 $P\{\underline{\theta}(X_1,X_2,\cdots X_n)<\theta<\overline{\theta}(X_1,X_2\cdots X_n)\}\geq 1-\alpha$，则称随机区间 $(\underline{\theta},\overline{\theta})$ 是 θ 的置信水平为 $1-\alpha$ 的置信区间，$1-\alpha$ 为置信水平

• 区间估计方法

ⓐ寻找一个样本 $X_1,X_2\cdots X_n$ 和 θ 的函数 $W=W(X_1,X_2,\cdots X_n;\theta)$，使得 W 的分布不依赖于 θ 以及其他未知参数，称具有这种性质的函数 W 为枢轴量

ⓑ对于给定的置信水平 $1-\alpha$，定出2个常数 a,b 使得 $P\{a<W(X_1,X_2,\cdots X_n;\theta)<b\}=1-\alpha$　（通常选 $a=-b$）

ⓒ不等式做等价变形，解出相应区间

	待估参量	其他参数	枢轴量的分布	置信区间
一个正态总体	μ	σ^2已知	$Z=\frac{\bar X-\mu}{\sigma/\sqrt n}\sim N(0,1)$	$\left(\bar X\pm\frac{\sigma}{\sqrt n}\cdot z_{\alpha/2}\right)$
	μ	σ^2未知	$t=\frac{\bar X-\mu}{S/\sqrt n}\sim t(n-1)$	$\left(\bar X\pm\frac{S}{\sqrt n}t_{\alpha/2}(n-1)\right)$
	σ^2	μ未知	$\chi^2=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$	$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)},\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right)$
两个正态总体	$\mu_1-\mu_2$	σ_1^2,σ_2^2已知	$Z=\frac{(\bar X-\bar Y)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)$	$\left(\bar X-\bar Y\pm z_{\alpha/2}\cdot\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$
	$\mu_1-\mu_2$	$\sigma_1^2=\sigma_2^2=\sigma^2$未知	$t=\frac{(\bar X-\bar Y)-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$ [其中 $S_w^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$]	$\left(\bar X-\bar Y\pm t_{\alpha/2}(n_1+n_2-2)\cdot S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$
	$\frac{\sigma_1^2}{\sigma_2^2}$	μ_1,μ_2未知	$F=\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$	$\left(\frac{S_1^2}{S_2^2}\cdot\frac{1}{F_{\alpha/2}(n_1-1,n_2-1)},\frac{S_1^2}{S_2^2}\cdot\frac{1}{F_{1-\alpha/2}(n_1-1,n_2-1)}\right)$

▲假设检验

①基本概念

• 显著性水平：若 $|z|=\left|\frac{\bar X-\mu_0}{\sigma/\sqrt n}\right|\geq z_{\alpha/2}$，则称 $\bar x$ 与 μ_0 的差异是显著的，则我们拒绝 H_0；反之，若 $|z|=\left|\frac{\bar X-\mu_0}{\sigma/\sqrt n}\right|<z_{\alpha/2}$，则称 $\bar x$ 与 μ_0 的差异是不显著的，则我们接受 H_0.　$P\{当H_0为真拒绝H_0\}=P_{\mu_0}\left\{\left|\frac{\bar X-\mu_0}{\sigma/\sqrt n}\right|\geq k\right\}=\alpha$　数 α 称为显著性水平

• 检验统计量：用于判断假设成立与否的统计量（统计量 $Z=\frac{\bar X-\mu_0}{\sigma/\sqrt n}$ 称检验统计量）

• 原假设备择假设：假设检验问题通常叙述为：在显著性水平 α 下，检验假设 $H_0:\mu=\mu_0$，$H_1:\mu\neq\mu_0$ 或称"在显著性水平 α 下，针对 H_1 检验 H_0"
H_0 称为原假设或零假设，H_1 称备择假设

• 两类错误：ⓐ当原假设 H_0 为真，观察值却落入拒绝域，而作出了拒绝 H_0 的判断，称弃真错误
ⓑ当原假设 H_0 不真，而观察值却落入接受域，而作出了接受 H_0 的判断，称取伪错误

	接受 H_0	拒绝 H_0
H_0 为真	正确	犯第一类错误
H_0 不真	犯第二类错误	正确

② 假设检验的一般步骤 衍玉

ⓐ 根据实际问题的要求，提出原假设 H_0 及备择假设 H_1；

ⓑ 在 H_0 为真的假设下寻找一个已知分布的统计量，它是样本的函数，不依赖于任何未知参数 (寻找统计量)

ⓒ 对于给定显著性水平 α 确定 H_0 的否定域，实质上是在假设 H_0 为真时寻找小概率事件：$P\{拒绝H_0 / H_0为真\} = \alpha$

ⓓ 由样本值算出检验统计量的值，看其是否在拒绝域中

ⓔ 做出判断是否接受还是拒绝 H_0

③ 检验统计量与拒绝域表格 (显著水平为 α)

	原假设	检验统计量	备择假设	拒绝域
1个总体	$\mu \le \mu_0$ $\mu \ge \mu_0$ $\mu = \mu_0$ (σ^2 已知)	$z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$	$\mu > \mu_0$ $\mu < \mu_0$ $\mu \ne \mu_0$	$z \ge z_\alpha$ $z \le -z_\alpha$ $\lvert z\rvert \ge z_{\alpha/2}$
[典例] P_{182} 例2	$\mu \le \mu_0$ $\mu \ge \mu_0$ $\mu = \mu_0$ (σ^2 未知)	$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$	$\mu > \mu_0$ $\mu < \mu_0$ $\mu \ne \mu_0$	$t \ge t_\alpha(n-1)$ $t \le -t_\alpha(n-1)$ $\lvert t\rvert \ge t_{\alpha/2}(n-1)$
	$\sigma^2 \le \sigma_0^2$ $\sigma^2 \ge \sigma_0^2$ $\sigma^2 = \sigma_0^2$ (μ 未知)	$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$	$\sigma^2 > \sigma_0^2$ $\sigma^2 < \sigma_0^2$ $\sigma^2 \ne \sigma_0^2$	$\chi^2 \ge \chi^2_\alpha(n-1)$ $\chi^2 \le \chi^2_{1-\alpha}(n-1)$ $\chi^2 \ge \chi^2_{\alpha/2}(n-1)$ 或 $\chi^2 \le \chi^2_{1-\alpha/2}(n-1)$
2个总体	$\mu_1 - \mu_2 \le \delta$ $\mu_1 - \mu_2 \ge \delta$ $\mu_1 - \mu_2 = \delta$ (σ_1^2, σ_2^2 已知)	$z = \frac{\bar{X} - \bar{Y} - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$	$\mu_1 - \mu_2 > \delta$ $\mu_1 - \mu_2 < \delta$ $\mu_1 - \mu_2 \ne \delta$	$z \ge z_\alpha$ $z \le z_\alpha$ $\lvert z\rvert \ge z_{\alpha/2}$
	$\mu_1 - \mu_2 \le \delta$ $\mu_1 - \mu_2 \ge \delta$ $\mu_1 - \mu_2 = \delta$ ($\sigma_1^2 = \sigma_2^2 = \sigma^2$ 未知)	$t = \frac{\bar{X} - \bar{Y} - \delta}{S_W\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ (其中 $S_W = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$)	$\mu_1 - \mu_2 > \delta$ $\mu_1 - \mu_2 < \delta$ $\mu_1 - \mu_2 \ne \delta$	$t \ge t_\alpha(n_1+n_2-2)$ $t \le t_\alpha(n_1+n_2-2)$ $\lvert t\rvert \ge t_{\alpha/2}(n_1+n_2-2)$
[典例] P_{185} 例2	$\sigma_1^2 \le \sigma_2^2$ $\sigma_1^2 \ge \sigma_2^2$ $\sigma_1^2 = \sigma_2^2$ (μ_1, μ_2 未知)	$F = \frac{S_1^2}{S_2^2}$	$\sigma_1^2 > \sigma_2^2$ $\sigma_1^2 < \sigma_2^2$ $\sigma_1^2 \ne \sigma_2^2$	$F \ge F_\alpha(n_1-1, n_2-1)$ $F \le F_{1-\alpha}(n_1-1, n_2-1)$ $F \ge F_{\alpha/2}(n_1-1, n_2-1)$ 或 $F \le F_{1-\alpha/2}(n_1-1, n_2-1)$
成对数据 [典例] P_{186} 例3	$\mu_D \le 0$ $\mu_D \ge 0$ $\mu_D = 0$	$t = \frac{\bar{D} - 0}{S_D/\sqrt{n}}$	$\mu_D > 0$ $\mu_D < 0$ $\mu_D \ne 0$	$t \ge t_\alpha(n-1)$ $t \le -t_\alpha(n-1)$ $\lvert t\rvert \ge t_{\alpha/2}(n-1)$