# 第一章 概率论的基本概念

1. 和事件: $A\cup B$. $A,B$至少发生一个.

积事件. $A\cap B$. $A,B$同时发生.

差事件: $A-B$. $A$发生, $B$不发生.

互斥: $A\cap B=\emptyset$

对立: $A\cup B=S$, $A\cap B=\emptyset$.

2. 德摩根律: $\overline{A\cup B}=\bar{A}\cap\bar{B}$, $\overline{A\cap B}=\bar{A}\cup\bar{B}$

交换、结合、分配律…

3. 加法公式. $P(A\cup B)=P(A)+P(B)-P(AB)$

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

4. 条件概率.

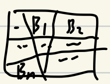
在$A$发生条件下发生$B$的概率 $P(B|A)=\dfrac{P(AB)}{P(A)}$

$\Rightarrow$乘法定理 $P(AB)=P(B|A)P(A)$ 多步事件.每一步的概率.

$$P(A_1A_2\cdots A_n)=P(A_n|A_1A_2\cdots A_{n-1})P(A_{n-1}|A_1A_2\cdots A_{n-2})\cdots P(A_2|A_1)P(A_1)$$

5. 全概率公式

设 $B_1,B_2\cdots B_n$为$S$的一个划分




$$P(A)=\sum_{i=1}^{n}P(A|B_i)P(B_i)$$

6. 贝叶斯公式.

常用形态: $P(B|A)=\dfrac{P(AB)}{P(A)}=\dfrac{P(A|B)P(B)}{P(A|B)P(B)+P(A|\bar{B})P(\bar{B})}$

一般形态: $P(B_i|A)=\dfrac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n}P(A|B_j)P(B_j)}$

7. 独立性 $P(AB)=P(A)P(B)$. $\Leftrightarrow$ 独立.

$$\left.\begin{aligned}P(AB)&=P(A)P(B)\\P(AC)&=P(A)P(C)\\P(BC)&=P(B)P(C)\\P(ABC)&=P(A)P(B)P(C)\end{aligned}\right\}\Leftrightarrow \text{独立.}$$

# 第二章 随机变量及其分布

1. 大写字母为随机变量.
   小写字母为实数.

2. 离散型随机变量分布律.

$$P(X=x_k)=p_k,\ k=1,2,3\cdots$$

或表格.

| $X$ | $x_1$ | $x_2$ | $\cdots$ | $x_n$ | $\cdots$ |
|---|---|---|---|---|---|
| $p_k$ | $p_1$ | $p_2$ | $\cdots$ | $p_n$ | $\cdots$ |

3. 0-1 分布.

| $X$ | 0 | 1 |
|---|---|---|
| $p_k$ | $1-p$ | $p$ |

$$P(X=k)=p^k(1-p)^{1-k},\ k=0,1.$$

4. 二项分布.

$X\sim b(n,p)$. $P(X=k)=C_n^k p^k q^{n-k},\ k=0,1,2\cdots$

$n$ 重伯努利试验

5. 泊松分布.

$$X\sim\pi(\lambda).\ P(X=k)=\frac{\lambda^k e^{-\lambda}}{k!}$$

泊松定理. $np_n=\lambda$

$$\lim_{n\to\infty}C_n^k p_n^k(1-p_n)^{n-k}=\frac{\lambda^k e^{-\lambda}}{k!}$$

当 $n$ 很大 $p_n$ 很小时, 二项分布趋于泊松分布

6. 随机变量分布函数.

$$F(x)=P(X\le x)\quad -\infty<x<+\infty.$$

7. 连续型随机变量.

$\exists f(t)$. s.t. $F(x)=\int_{-\infty}^{x}f(t)\,dt$. $\quad \frac{dF(x)}{dx}=f(x)$

$f(x)$ 为 $X$ 的概率密度函数.

具有归一性.

8. 均匀分布.

$$X\sim U(a,b)\qquad f(x)=\begin{cases}\frac{1}{b-a}, & a<x<b\\ 0, & 其它.\end{cases}\qquad F(x)=\begin{cases}0, & x<a.\\ \frac{x-a}{b-a}, & a\le x<b\\ 1, & x\ge b\end{cases}$$

9. 指数分布.

$$f(x)=\begin{cases}\frac{1}{\theta}e^{-\frac{x}{\theta}}, & x>0\\ 0, & 其它.\end{cases}$$ 参数为θ的指数分布.

$$F(x)=\begin{cases}1-e^{-\frac{x}{\theta}}, & x>0.\\ 0, & 其它.\end{cases}$$

10. 正态分布. (高斯分布).

$X\sim N(\mu,\sigma^2)$. $f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$, $-\infty<x<\infty$.

$\int_{-\infty}^{+\infty}e^{-\frac{t^2}{2}}dt=\sqrt{2\pi}$. (证明：平方化重积分后 极坐标换元)

☆ $Y=\frac{X-\mu}{\sigma}\sim N(0,1)$

标准正态分布. $X\sim N(0,1)$. $\varphi(x)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt.$$

$$\Phi(-x)=1-\Phi(x)$$

11. 随机变量的函数的分布.

$f_X(x)$, $-\infty<x<+\infty$, $Y=g(X)$ 为连续~~严格~~单调函数. 可得 $X=h(Y)$

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

$\alpha=\min(g(-\infty),g(\infty))$, $\beta=\max(g(-\infty),g(\infty))$

具体操作：在 $F_Y(y)$ 中. $Y\le y\Rightarrow g(X)\le y$ 解出 $X$.

如 $X\le h(Y)$

得到 $F_X(h(Y))$

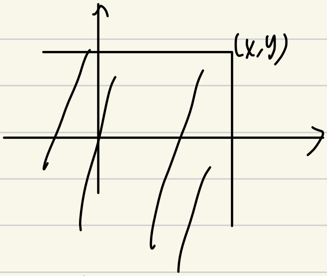
则 $f_Y(y)=\frac{dF_X(h(y))}{dy}$

$=f_X(h(y))\cdot|h'(y)|$

# 第三章、多维随机变量及其分布.

## 1. 二维随机变量的联合分布函数.

$$F(x,y)=P(X\le x,\ Y\le y)$$




$$P(x_1<X\le x_2,\ y_1<Y\le y_2)$$
$$=F(x_2,y_2)-F(x_2,y_1)-F(x_1,y_2)+F(x_1,y_1)$$

## 2. 二维离散型随机变量的联合分布律

$$P(X=x_i,\ Y=y_j)=p_{ij},\quad i,j=1,2\cdots$$

也可为表格

| $Y \backslash X$ | $x_1$ | $x_2$ | $\cdots$ | $x_i$ | $\cdots$ |
|---|---|---|---|---|---|
| $y_1$ | $p_{11}$ | $p_{21}$ | $\cdots$ | $p_{i1}$ | $\cdots$ |
| $y_2$ | $p_{12}$ | $p_{22}$ | $\cdots$ | $p_{i2}$ | $\cdots$ |
| $\vdots$ | $\vdots$ | $\vdots$ | | $\vdots$ | |
| $y_j$ | $p_{1j}$ | $p_{2j}$ | $\cdots$ | $p_{ij}$ | $\cdots$ |
| $\vdots$ | $\vdots$ | $\vdots$ | | | |

## 3. 二维连续型随机变量

$$F(x,y)=\int_{-\infty}^{y}\int_{-\infty}^{x} f(u,v)\,du\,dv.$$

$f(u,v)$ 为联合概率密度

$$\frac{\partial^2 F(x,y)}{\partial x\partial y}=f(x,y)$$

## 4. 边缘分布函数.

$$F_X(x)=F(X\le x, Y<\infty)=F(x,\infty)$$
$$F_Y(y)=F(\infty,y)$$

离散型: $F_X(x)=\sum\limits_{x_i\le x}\sum\limits_{j=1}^{\infty} p_{ij}.$

X的边缘分布律 $P(X=x_i)=\sum\limits_{j=1}^{\infty} p_{ij}=p_{i\cdot}$

Y的边缘分布律. $P(Y=y_j)=\sum\limits_{i=1}^{\infty} p_{ij}=p_{\cdot j}.$

连续型：$F(x)=F(x,\infty)=\int_{-\infty}^{x}\left(\int_{-\infty}^{+\infty}f(x,y)\,dy\right)dx$

$\Rightarrow$ X的边缘概率密度 $f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\,dy$

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

5. 二维正态分布.

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left[-\frac{1}{2}\cdot\frac{1}{(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right]$$

$\rho\in(-1,1)$

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

$$f_X(x)=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\qquad f_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}$$

二维正态的两个边缘分布都是一维正态分布.

6. 二维离散型条件分布.

$$P(X=x_i|Y=y_j)=\frac{P(X=x_i,Y=y_j)}{P(Y=y_j)}=\frac{p_{ij}}{p_{\cdot j}}$$

$$P(Y=y_j|X=x_i)=\frac{p_{ij}}{p_{i\cdot}}$$

7. 二维连续型条件分布.

条件概率密度. $f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}$

$P(X\le x|Y=y)=F_{X|Y}(x|y)=\int_{-\infty}^{x}f_{X|Y}(x|y)\,dx=\int_{-\infty}^{x}\frac{f(x,y)}{f_Y(y)}dx$ 为在$Y=y$条件下X的条件分布函数.

8. 相互独立的随机变量.

连续：$F(x,y)=F_X(x)F_Y(y)$

或 $f(x,y)=f_X(x)f_Y(y)$

离散：$P(X=x_i,Y=y_j)=P(X=x_i)P(Y=y_j)$

9. 正态随机变量独立的充要条件.

$$\rho = 0$$

10. 两个随机变量的函数的分布.

①. $Z = X + Y$.

$$f_{X+Y}(z) = \int_{-\infty}^{+\infty} f(z-y, y)\, dy$$

$$= \int_{-\infty}^{+\infty} f(x, z-x)\, dx$$

若 $X, Y$ 独立.

$$f_{X+Y}(z) = \int_{-\infty}^{+\infty} f_X(z-y) f_Y(y)\, dy$$

$$= \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x)\, dx.$$

卷积公式.

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.

②. $Z = \frac{Y}{X}$.

$$f_{\frac{Y}{X}}(z) = \int_{-\infty}^{+\infty} |x| f(x, xz)\, dx \overset{XY\text{独立}}{=\!=\!=} \int_{-\infty}^{+\infty} |x| f_X(x) f_Y(xz)\, dx$$

③. $Z = XY$

$$f_{XY}(z) = \int_{-\infty}^{+\infty} \frac{1}{|x|} f(x, \frac{z}{x})\, dx \overset{XY\text{独立}}{=\!=\!=} \int_{-\infty}^{+\infty} \frac{1}{|x|} f_X(x) f_Y(\frac{z}{x})\, dx.$$

④. $Z = \max\{X, Y\}$, $\min\{X, Y\}$

$$F_{max}(z) = P(X \le z) P(Y \le z)$$

$$= F_X(z) F_Y(z)$$

$$F_{min}(z) = 1 - (1 - F_X(z))(1 - F_Y(z))$$

第四章. 随机变量的数字特征.

1. 期望. 均值.

$$\begin{cases} \text{离散.} \quad E(X)=\sum_{k=1}^{\infty} x_k p_k \quad (\text{若级数绝对收敛}) \\ \text{连续.} \quad E(X)=\int_{-\infty}^{+\infty} x f(x)\,dx. \end{cases}$$

若条件收敛,可能得到不同期望

$Y=g(X)$

$$E(Y)=E(g(X))=\begin{cases} \sum_{k=1}^{\infty} g(x_k)p_k \\ \int_{-\infty}^{+\infty} g(x) f(x)\,dx. \end{cases}$$

$Z=g(X,Y)$

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

2. 期望性质

$$\begin{cases} E(c)=C \\ E(cX)=CE(X) \\ E(X+Y)=E(X)+E(Y) \\ X,Y\text{独立} \Rightarrow E(XY)=E(X)E(Y) \end{cases}$$

3. 方差.

$$D(X)=Var(X)=E((X-E(X))^2)$$

标准差. $\sigma(X)=\sqrt{D(X)}$

$$\begin{cases} \text{离散} \quad D(X)=\sum_{k=1}^{\infty} (x_k-E(X))^2 p_k \\ \text{连续} \quad D(X)=\int_{-\infty}^{+\infty} (x-E(X))^2 f(x)\,dx. \end{cases}$$

$$D(X)=E(X^2)-(E(X))^2$$

4. 泊松分布的 $\begin{cases} \text{期望.} & \lambda \\ \text{方差.} & \lambda \end{cases}$

5. 均匀分布的 $\begin{cases} \text{期望} & \frac{a+b}{2} \\ \text{方差} & \frac{(b-a)^2}{12} \end{cases}$

6. 指数分布的 $\begin{cases} \text{期望} & \theta \\ \text{方差} & \theta^2 \end{cases}$

7. 正态分布的 $\begin{cases} \text{期望} & \mu \\ \text{方差} & \sigma^2 \end{cases}$

8. 二项分布 $\begin{cases} E(X) = np \\ D(X) = np(1-p) \end{cases}$

9. 超几何分布. $E(X) = n \cdot \frac{M}{N}$ . n次, 从N中抽M个.

10. 方差性质

① $D(C) = 0$.

②. $D(CX) = C^2 D(X)$

③. $D(X+C) = D(X)$

④. $D(X+Y) = \begin{cases} D(X) + D(Y) + 2E\left[(X-E(X))(Y-E(Y))\right] = D(X)+D(Y)+2Cov(X,Y) \\ \text{独立. } D(X)+D(Y) \end{cases}$

⑤. $D(X) = 0 \iff P(X = E(X)) = 1$.

11. 切比雪夫不等式

$$P(|X-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

随机变量偏离期望的概率小于一个值. (粗略估计)

12. 协方差.

$$Cov(X,Y) = E\left[(X-E(X))(Y-E(Y))\right]$$

$$D(X+Y) = D(X) + D(Y) + 2Cov(X,Y)$$

13. 协方差性质

$Cov(X,Y) = Cov(Y,X)$

$Cov(X,X) = D(X)$

☆ $Cov(X,Y) = E(XY) - E(X)E(Y)$

$Cov(aX,bY) = ab\,Cov(X,Y)$

$Cov(X_1+X_2,Y) = Cov(X_1,Y) + Cov(X_2,Y)$

14. 相关系数

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}} \in [-1,1]$$

$$|\rho_{XY}| = 1 \iff Y = a + bX.$$

⚠ $\rho_{XY} = 0$时, X和Y不线性相关, 但不代表相互独立.

15. 矩.

k阶原点矩 $E(X^k)$

k阶中心矩 $E\left[(X-E(X))^k\right]$

# 第五章. 大数定律和中心极限定理.

## 1. 弱大数定律.

$X_1 \cdots X_n$ 独立同分布. $E(X_k)=\mu$. $\forall \varepsilon > 0$.

则 $\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| < \varepsilon\right) = 1$

序列 $\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \mu$. 依概率收敛.

## 2. 伯努利大数定律.

独立重复试验, $f_A$ 为 A 发生次数 ( n次试验 ), $\forall \varepsilon > 0$.

$$\begin{cases} \lim_{n\to\infty} P\left(\left|\frac{f_A}{n} - p\right| < \varepsilon\right) = 1 \\ \lim_{n\to\infty} P\left(\left|\frac{f_A}{n} - p\right| \geq \varepsilon\right) = 0. \end{cases}$$

频率趋向于概率.

## 3. 中心极限定理.

$X_1 \cdots X_n$ 独立同分布. $E(X_k)=\mu$. $D(X_k)=\sigma^2$, $n\to\infty$ 时.

$$Y_n = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\,\sigma} \overset{\text{近似}}{\sim} N(0,1)$$

$$= \frac{\frac{1}{n}\sum_{k=1}^{n} X_k - \mu}{\frac{\sigma}{\sqrt{n}}}$$

即 $\frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \overset{\text{近似}}{\sim} N(0,1)$

$\overline{X} \overset{\text{近似}}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$

## 4. D-L 定理

$\eta_n \sim b(n,p)$. , $n\to\infty$ 时. $\forall \varepsilon > 0$.

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

二项分布的极限是正态分布.

第六章 样本及抽样分布

1. 样本，总体

$X$ 为1随机变量

$X_1 \cdots X_n$ 也是，同分布，称这些随机变量为样本

$x_1 \cdots x_n$ 是样本观察值

总体是试验全部可能值

2. 统计量

统计量 $g(X_1, \cdots X_n)$ 是关于 $(X_1 \cdots X_n)$ 的函数，不含任何未知参数

一些重要统计量

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

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

样本标准差 $S = \sqrt{S^2}$

样本k阶原点矩 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$　　$A_1 = \bar{X}$，$A_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$

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

3. $\chi^2$ 分布

$\chi^2 = X_1^2 + X_2^2 + \cdots X_n^2$，$X_1 \cdots X_n$ 来自 $N(0,1)$

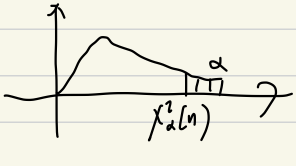
$\chi^2 \sim \chi^2(n)$

性质：可加性　$\chi^2(n_1) + \chi^2(n_2) = \chi^2(n_1 + n_2)$

$E(\chi^2) = n$　　$D(\chi^2) = 2n$

上分位数 $P(\chi^2 > \chi^2_\alpha(n)) = \alpha$

$n \to \infty$ 时，$\chi^2_\alpha(n) = \frac{1}{2}(z_\alpha + \sqrt{2n-1})^2$




4. t分布

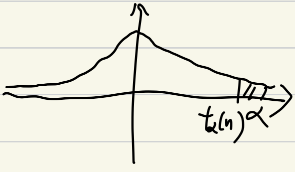
$X \sim N(0,1)$，$Y \sim \chi^2(n)$

$$t = \frac{X}{\sqrt{\frac{Y}{n}}} \sim t(n)$$

对称分布，关于 $x=0$

上α分位数，$P(t > t_\alpha(n)) = \alpha$

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



## 5. F分布.

$X \sim \chi^2(n_1)$, $Y \sim \chi^2(n_2)$

$$F = \frac{X/n_1}{Y/n_2} \sim F(n_1, n_2)$$


上α分位数. $P(F > F_\alpha(n_1,n_2)) = \alpha$

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

## 6. 一些性质

① $E(\bar{X}) = \mu$

$D(\bar{X}) = \frac{\sigma^2}{n}$.

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

$E(X^2) = D(X) + E^2(X)$

$E(S^2) = \sigma^2$

$E(S_n^2) = \frac{n}{n-1}\sigma^2$

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

$= \frac{1}{n-1}\left(n(\sigma^2+\mu^2) - n\left(\frac{\sigma^2}{n}+\mu^2\right)\right)$

$= \frac{1}{n-1}\left((n-1)\sigma^2\right)$

$= \sigma^2$.

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

证: $E(S^2) = \sigma^2$ $D(S^2) = \frac{2\sigma^4}{n-1}$

$D\left[\frac{(n-1)S^2}{\sigma^2}\right] = \frac{(n-1)^2}{\sigma^4} \cdot \frac{2\sigma^4}{n-1} = 2(n-1)$

$E\left(\frac{n-1}{\sigma^2}S^2\right) = n-1$.

$\Rightarrow \chi^2(n-1)$

③. $\bar{X}$ 与 $S^2$ 相互独立.

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

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

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

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

对于2个样本 $X_1 \cdots X_{n_1}$  $Y_1 \cdots Y_{n_2}$.

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

⑤. $\frac{\frac{S_1^2}{\sigma_1^2}}{\frac{S_2^2}{\sigma_2^2}} \sim F(n_1-1, n_2-1)$

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

⑥. 当 $\sigma_1^2 = \sigma_2^2 = \sigma^2$

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

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

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

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

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

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

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

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

# 第七章. 参数估计.

## (I). 点估计.

### 1. 矩估计.

$f(x, \theta_1 \cdots \theta_k)$  $\theta_i$ 为待估参数

总体矩 $\mu_l = E(X^l) = \int_{-\infty}^{+\infty} x^l f(x, \theta_1 \cdots \theta_k)\, dx$.

样本矩 $A_l = \frac{1}{n}\sum_{i=1}^{n} X_i^l \xrightarrow{P} \mu_l$.

$$\begin{cases} \mu_1 = \mu_1(\theta_1 \cdots \theta_k) \\ \vdots \\ \mu_k = \mu_k(\theta_1 \cdots \theta_k) \end{cases} \Rightarrow \begin{cases} \theta_1 = \theta_1(\mu_1 \cdots \mu_k) \\ \vdots \\ \theta_k = \theta_k(\mu_1 \cdots \mu_k) \end{cases} \Rightarrow \begin{cases} \hat{\theta}_1 = \theta_1(A_1 \cdots A_k) \\ \vdots \\ \hat{\theta}_k = \theta_k(A_1 \cdots A_k) \end{cases}$$

## 2. 最大似然估计.

$X_1 \cdots X_n$ 是样本. 独立同分布.

① 离散. $P(X=x) = p(x,\theta)$

$$L(\theta) = \prod_{i=1}^{n} P(x_i,\theta) = L(x_1 \cdots x_n, \theta)$$

② 连续. $\forall X_i$. $f(x,\theta)$.

$$L(\theta) = \prod_{i=1}^{n} f(x_i,\theta) = L(x_1 \cdots x_n, \theta)$$

$L(\theta)$ 称似然函数.

$$L(x_1 \cdots x_n, \hat{\theta}) = \max_{\theta \in \Theta} L(x_1, x_2 \cdots x_n, \theta)$$

$\hat{\theta}(x_1 \cdots x_n)$ 称 $\theta$ 的最大似然估计值, $\hat{\theta}(X_1, \cdots X_n)$ 称最大似然估计量.

求解 $\hat{\theta}$. $\frac{d}{d\theta} \ln L(\theta) = 0$ 对数似然方程.

多个未知参数 $\theta_1 \cdots \theta_k$.

令 $\frac{\partial}{\partial \theta_k} \ln L = 0$. 解出 $\hat{\theta}_i$.

例. $X \sim U(a,b)$, $a,b$ 待估.

$X_{(1)} = \min(X_1 \cdots X_n)$, $X_{(n)} = \max(X_1 \cdots X_n)$

$$f(x,a,b) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0, & \text{其他} \end{cases}$$

$$L(a,b) = \frac{1}{(b-a)^n}$$

$X_{(1)} \ge a$. $X_{(n)} \le b$.

要使 $L$ 最大. $(b-a)$ 最小.

$$\Rightarrow L(a,b) = \frac{1}{(b-a)^n} \le \frac{1}{(X_{(n)} - X_{(1)})^n}$$

$\Rightarrow \hat{a} = \min_{1 \le i \le n} x_i$, $\hat{b} = \max_{1 \le i \le n} x_i$ 为最大似然估计值.

估计量的评价标准.

3. 无偏性.　　　　纠偏：$E(\hat{\theta})=a\theta+b$. $\frac{1}{a}(\hat{\theta}-b)$ 无偏

$$E(\hat{\theta})=\theta.$$

4. 有效性.

$$D(\hat{\theta}_1)\le D(\hat{\theta}_2)$$ $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 更有效.

5. 相合性.(基本要求)

$\forall \varepsilon>0$, $\lim\limits_{n\to\infty} P(|\hat{\theta}-\theta|<\varepsilon)=1$.

例：$S^2$ 是 $\sigma^2$ 的无偏、相合估计量

$S_n^2$ 是 $\sigma^2$ 的有偏、相合估计量.

(II). 区间估计.

1. 置信区间.

置信下限　置信上限

未知参数 $\theta$. 对于一个样本.有 $\theta$ 的2个统计量 $\underline{\theta}$ 和 $\overline{\theta}$ 若给定一个置信水平 $1-\alpha$ ($\alpha\in(0,1)$), $\theta$ 真值介于 $[\underline{\theta},\overline{\theta}]$ 的概率不小于 $1-\alpha$.

$$P\left(\underline{\theta}(X_1\cdots X_n)<\theta<\overline{\theta}(X_1\cdots X_n)\right)\ge 1-\alpha$$

这个区间包含真值的概率是 $1-\alpha$. $\alpha$ 是小量.

2. 寻找未知参数置信区间的一般方法.

①. 寻找枢轴量 $W(X_1,X_2\cdots X_n,\theta)$, $W$ 分布不依赖 $\theta$ 及其它参数.

②. 对于置信水平 $1-\alpha$. 寻找常数 $a,b$. 使 $P(a<W<b)=1-\alpha$.

解得 $(\underline{\theta},\overline{\theta})$ 为 $\theta$ 的一个 $(1-\alpha)$ 置信区间.

3. 单侧置信区间

$$\begin{cases} P(\theta\ge\underline{\theta})=1-\alpha & [\underline{\theta},+\infty) \\ P(\theta\le\overline{\theta})=1-\alpha. & (-\infty,\overline{\theta}) \end{cases}$$

是 $\theta$ 的单侧置信区间. $\dfrac{\underline{\theta}}{\overline{\theta}}$ 称为 $\theta$ 的单侧置信 $\dfrac{\text{下限}}{\text{上限}}$.

## 4. 正态总体均值与方差区间估计.

### ①. 单个总体 省流表格.

| 问题 | 置信区间 | 推导思路. |
|---|---|---|
| 已知 $\sigma^2$，求 $\mu$ 置信区间 | $\bar X \pm \frac{\sigma}{\sqrt n} z_{\frac{\alpha}{2}}$ | $W=\frac{\bar X-\mu}{\sigma/\sqrt n}\sim N(0,1)$, $P(-z_{\frac{\alpha}{2}}<W<z_{\frac{\alpha}{2}})=1-\alpha$. |
| 未知 $\sigma^2$，求 $\mu$ 置信区间 | $\bar X \pm \frac{S}{\sqrt n} t_{\frac{\alpha}{2}}(n-1)$ | $W=\frac{\bar X-\mu}{S/\sqrt n}\sim t(n-1)$, $P(-t_{\frac{\alpha}{2}}(n-1)<W<t_{\frac{\alpha}{2}}(n-1))=1-\alpha$. |
| 未知 $\mu$，求 $\sigma^2$ 置信区间<br>$\sigma$置信区间直接开根号 | $\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}, \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right)$ | $W=\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$, $P(\chi^2_{1-\frac{\alpha}{2}}(n-1)<W<\chi^2_{\frac{\alpha}{2}}(n-1))=1-\alpha$ |
| 无用. 已知 $\sigma^2$，求 $\mu$ 置信区间. | $\left[\frac{\sum_{i=1}^n(X_i-\mu)^2}{\chi^2_{\frac{\alpha}{2}}(n)}, \frac{\sum_{i=1}^n(X_i-\mu)^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right]$ | $W=\sum_{i=1}^n\left(\frac{X_i-\mu}{\sigma}\right)^2\sim\chi^2(n)$, $P(\chi^2_{1-\frac{\alpha}{2}}(n)<W<\chi^2_{\frac{\alpha}{2}}(n))=1-\alpha$ |

单侧置信上下限(先上限后下限)

$$\bar X\pm\frac{\sigma}{\sqrt n}z_\alpha$$

$$\bar X\pm\frac{S}{\sqrt n}t_\alpha(n-1)$$

$$\frac{(n-1)S^2}{\chi^2_{1-\alpha}(n-1)} \text{ 和 } \frac{(n-1)S^2}{\chi^2_{\alpha}(n-1)}$$

$$\frac{\sum_{i=1}^n(X_i-\mu)^2}{\chi^2_{1-\alpha}(n-1)} \text{ 和 } \frac{\sum_{i=1}^n(X_i-\mu)^2}{\chi^2_{\alpha}(n)}$$

### ②. 2个总体.

| 问题. | 置信区间 | 推导思路. |
|---|---|---|
| 已知 $\sigma_1^2$、$\sigma_2^2$，求 $\mu_1-\mu_2$ | $\left(\bar X-\bar Y\pm z_{\frac{\alpha}{2}}\cdot\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$ | $\frac{(\bar X-\bar Y)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}\sim N(0,1)$ |
| $\sigma_1^2=\sigma_2^2=\sigma^2$ 未知，求 $\mu_1-\mu_2$. | $\left(\bar X-\bar Y\pm t_{\frac{\alpha}{2}}(n_1+n_2-2)\cdot S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$ | $\frac{(\bar X-\bar Y)-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim t(n_1+n_2-2)$ |
| $\sigma_1^2,\sigma_2^2$ 未知. $n_1,n_2$ 较大. 求 $\mu_1-\mu_2$ | $\left(\bar X-\bar Y\pm z_{\frac{\alpha}{2}}\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\right)$ | $\frac{(\bar X-\bar Y)-(\mu_1-\mu_2)}{\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}}\sim N(0,1)$ |
| 无用 $\mu_1,\mu_2$ 已知，求 $\frac{\sigma_1^2}{\sigma_2^2}$ | $\left(\frac{B_{2X}}{B_{2Y}}\cdot\frac{1}{F_{\frac{\alpha}{2}}(n_1,n_2)}, \frac{B_{2X}}{B_{2Y}}\cdot\frac{1}{F_{1-\frac{\alpha}{2}}(n_1,n_2)}\right)$ | $\frac{\sum(X_i-\mu_1)^2}{n_1\sigma_1^2}\Big/\frac{\sum(Y_i-\mu_2)^2}{n_2\sigma_2^2}\sim F(n_1,n_2)$ |
| $\mu_1,\mu_2$ 未知，求 $\frac{\sigma_1^2}{\sigma_2^2}$ | $\left(\frac{S_1^2}{S_2^2}\cdot\frac{1}{F_{\frac{\alpha}{2}}(n_1-1,n_2-1)}, \frac{S_1^2}{S_2^2}\cdot\frac{1}{F_{1-\frac{\alpha}{2}}(n_1-1,n_2-1)}\right)$ | $\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$ |

$$S_w=\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}$$

单侧置信限(先上后下)

$$\bar X-\bar Y\pm z_\alpha\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}.$$

$$\bar X-\bar Y\pm t_\alpha(n_1+n_2-2)S_w\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}.$$

$$\bar X-\bar Y\pm z_\alpha\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}$$

$$\frac{B_{2X}}{B_{2Y}}\frac{1}{F_{1-\alpha}(n_1,n_2)} \text{ 和 } \frac{B_{2X}}{B_{2Y}}\frac{1}{F_{\alpha}(n_1,n_2)}$$

$$\frac{S_1^2}{S_2^2}\frac{1}{F_{1-\alpha}(n_1-1,n_2-1)} \text{ 和 } \frac{S_1^2}{S_2^2}\frac{1}{F_{\alpha}(n_1-1,n_2-1)}$$

# 第八章. 假设检验

1. 假设检验基本思路.

原假设 $H_0$

备择. $H_1$

假设 $H_0$. 有[illegible].

| | 接受 $H_0$ | 拒绝 $H_0$ |
|---|---|---|
| $H_0$为真 | ✓ | 第一类错误 弃真 |
| $H_0$为假 | 第二类错误 取伪 | ✓ |

$P$ ( $H_0$为真时拒绝$H_0$ ) $\le \alpha$.

2. 正态总体均值、方差检验法. (显著性水平为$\alpha$)

规律：单边$\alpha$. 双边$\frac{\alpha}{2}$.

| 原假设 $H_0$ | 检验统计量 | 拒绝域 |
|---|---|---|
| $\sigma^2$已知时, $\mu \le \mu_0$ <br> $\mu \ge \mu_0$ <br> $\mu = \mu_0$ | $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ | $Z \ge Z_\alpha$ <br> $Z \le -Z_\alpha$ <br> $\lvert Z \rvert \ge Z_{\frac{\alpha}{2}}$. |
| $\sigma^2$未知时 $\mu \le \mu_0$ <br> $\mu \ge \mu_0$ <br> $\mu = \mu_0$ | $t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$ | $t \ge t_\alpha(n-1)$ <br> $t \le -t_\alpha(n-1)$ <br> $\lvert t \rvert \ge t_{\frac{\alpha}{2}}(n-1)$ |
| $\sigma_1^2, \sigma_2^2$ 已知 $\mu_1 - \mu_2 \le \delta$ <br> $\mu_1 - \mu_2 \ge \delta$ <br> $\mu_1 - \mu_2 = \delta$ | $Z = \frac{\bar{X} - \bar{Y} - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$ | $Z \ge Z_\alpha$ <br> $Z \le -Z_\alpha$ <br> $\lvert Z \rvert \ge Z_{\frac{\alpha}{2}}$. |
| $\sigma_1^2 = \sigma_2^2 = \sigma^2$未知 $\mu_1 - \mu_2 \le \delta$ <br> $\mu_1 - \mu_2 \ge \delta$ <br> $\mu_1 - \mu_2 = \delta$ | $t = \frac{\bar{X} - \bar{Y} - \delta}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$, $S_w = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}}$ | $t \ge t_\alpha(n_1 + n_2 - 2)$ <br> $t \le -t_\alpha(n_1 + n_2 - 2)$ <br> $\lvert t \rvert \ge t_{\frac{\alpha}{2}}(n_1 + n_2 - 2)$ |
| $\mu$未知. $\sigma \le \sigma_0^2$ <br> $\sigma^2 \ge \sigma_0^2$ <br> $\sigma^2 = \sigma_0^2$ | $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$ | $\chi^2 \ge \chi^2_\alpha(n-1)$ <br> $\chi^2 \le \chi^2_{1-\alpha}(n-1)$ <br> $\chi^2 \ge \chi^2_{\frac{\alpha}{2}}(n-1)$ 或 $\chi^2 \le \chi^2_{1-\frac{\alpha}{2}}(n-1)$ |

| | | | |
|---|---|---|---|
| $u_1, u_2$ 未知. | $\sigma_1^2 \leq \sigma_2^2$ | $F = \frac{S_1^2}{S_2^2}$ | $F \geq F_\alpha(n_1-1, n_2-1)$ |
| | $\sigma_1^2 \geq \sigma_2^2$ | | $F \leq F_{1-\alpha}(n_1-1, n_2-1)$ |
| | $\sigma_1^2 = \sigma_2^2$ | | $F \geq F_{\frac{\alpha}{2}}(n_1-1, n_2-1)$ 或 $F \leq F_{1-\frac{\alpha}{2}}(n_1-1, n_2-1)$ |
| 成对数据检验 | $u_D \leq 0$ | $t = \frac{\bar{D} - 0}{S_D / \sqrt{n}}$ | $t \geq t_\alpha(n-1)$ |
| | $u_D \geq 0$ | | $t \leq -t_\alpha(n-1)$ |
| | $u_D = 0$. | | $\lvert t \rvert \geq t_{\frac{\alpha}{2}}(n-1)$ |
| $(D_i = X_i - Y_i)$ $D_1 \cdots D_n$ 构成正态总体样本 | | | |

第九章. 方差分析及回归分析.

1. 单因素试验的方差分析的数学模型.

假定因素A的s个水平. $A_1 \cdots A_s$，对于 $A_j$，有样本 $X_{1j}, X_{2j} \cdots X_{n_j j}$，来自具有相同方差 $\sigma^2$，均值分别为 $\mu_j$ 的正态总体 $N(\mu_j, \sigma^2)$，$\mu_j, \sigma^2$ 未知. 设不同的样本相互独立.

$X_{ij} \sim N(\mu_j, \sigma^2)$，则 $\varepsilon_{ij} = X_{ij} - \mu_j \sim N(0, \sigma^2)$ 为随机误差.

则
$$\begin{cases} X_{ij} = \mu_j + \varepsilon_{ij} \\ \varepsilon_{ij} \sim N(0, \sigma^2), \text{各 } \varepsilon_{ij} \text{独立.} \\ i = 1, 2, \cdots n_j, \quad j = 1, 2 \cdots s \end{cases}$$
$\mu_j$ 和 $\sigma^2$ 均为未知参数.

任务：① 检验 s个总体 $N(\mu_1, \sigma^2) \cdots N(\mu_s, \sigma^2)$ 均值是否相等.

$H_0: \mu_1 = \mu_2 = \cdots = \mu_s$

$H_1: \mu_1 \cdots \mu_s$ 不全相等

② 作出 $\mu_1, \mu_2 \cdots \mu_s, \sigma^2$ 的估计.

另一形式：记 $\mu = \frac{1}{n}\sum_{j=1}^{s} n_j \mu_j$. $n = \sum_{j=1}^{s} n_j$

总平均.

$\delta_j = \mu_j - \mu$. 则 $n_1\delta_1 + n_2\delta_2 + \cdots n_s\delta_s = 0$.

$\delta_j$ 表示水平 $A_j$ 下总体均值与总平均的差异，又叫 $A_j$ 的效应.

模型：
$$\begin{cases} X_{ij} = \mu + \delta_j + \varepsilon_{ij} \\ \varepsilon_{ij} \sim N(0, \sigma^2), \ \varepsilon_{ij} \text{ 独立.} \\ i = 1, 2 \cdots n_j, \quad j = 1, 2 \cdots s \\ \sum_{j=1}^{s} n_j \delta_j = 0. \end{cases}$$

$H_0: \delta_1 = \delta_2 = \cdots = \delta_s = 0$

$H_1: \delta_1 \cdots \delta_s$ 不全为 0.

2. 平方和分解.

总偏差平方和. $S_T = \sum_{j=1}^{s}\sum_{i=1}^{n_j}(X_{ij} - \bar{X})^2$，$\bar{X} = \frac{1}{n}\sum_{j=1}^{s}\sum_{i=1}^{n_j} X_{ij}$ 为总平均.

设 $\bar{X}_{\cdot j} = \frac{1}{n_j}\sum_{i=1}^{n_j} X_{ij}$

可证明 $S_T = S_E + S_A$，$S_E = \sum_{j=1}^{s}\sum_{i=1}^{n_j}(X_{ij} - \bar{X}_{\cdot j})^2$，

$S_A = \sum_{j=1}^{s}\sum_{i=1}^{n_j}(\bar{X}_{\cdot j} - \bar{X})^2 = \sum_{j=1}^{s} n_j(\bar{X}_{\cdot j} - \bar{X})^2 = \sum_{j=1}^{s} n_j \bar{X}_{\cdot j}^2 - n\bar{X}^2$

$S_E$ 各项 $(X_{ij}-\bar{X}_{\cdot j})^2$ 表示在水平 $A_j$ 下，样本观察值与样本均值差异，称为误差平方和

$S_A$ 和项 $n_j(\bar{X}_{\cdot j}-\bar{X})^2$ 表示在水平 $A_j$ 下，样本均值与总平均的差异，称为效应平方和

3. $S_E$, $S_A$ 的统计特性.

①. $S_E$：$\frac{S_E}{\sigma^2} \sim \chi^2(n-s)$，$n=\sum_{j=1}^{s} n_j$

[证：$S_E=\sum_{j=1}^{s}\sum_{i=1}^{n_j}(X_{ij}-\bar{X}_{\cdot j})^2$，$\frac{\sum_{i=1}^{n_j}(X_{ij}-\bar{X}_{\cdot j})^2}{\sigma^2} \sim \chi^2(n_j-1)$)

$E(S_E)=(n-s)\sigma^2$.

②. $S_A$：　$E(S_A)=(s-1)\sigma^2+\sum_{j=1}^{s} n_j\delta_j^2$

$S_A$ 与 $S_E$ 独立，当 $H_0$ 为真时，$\frac{S_A}{\sigma^2} \sim \chi^2(s-1)$

4. 假设检验问题的拒绝域.

第9章仅考单因子方差分析.