概率论：研究随机现象规律性.

随机试验：①可重复 ②可能结果不止一个,能事先明确所有结果
③进行试验前结果不确定

样本空间 ⇐ 样本点

随机事件 → 事件的运算：和(并). 积(交). 差
互斥. 对立. 包含

频率 → 概率 ⇒ 运算：非负. 规范. 可列可加

古典概型：等可能.

几何概型：等可能 & 几何区间

条件概率：$P(B|A)=\frac{P(AB)}{P(A)}$

$P(AB)=P(A)P(B|A)$

全概率公式

贝叶斯公式

随机变量：样本空间 → R 的一个映射 { 离散 / 连续

↓

概率分布 → 分布函数：$F(x)=P(X\le x)$. 记为 $X\sim F(x)$

单调. 有界. 右连续

分布列(离散型)

对连续型：概率密度函数 $p(x)$

$F(x)=\int_{-\infty}^{x}p(t)\,dt$

连续型分布函数一定连续但密度函数必是非负可积

期望：
- 离散型 $\sum_{i=1}^{n}x_iP(x_i)$ 需绝对收敛！
- 连续型 $\int_{-\infty}^{\infty}xf(x)\,dx$

运算律

常用离散型分布

| | 概率分布 | 期望 | 方差 | |
|---|---|---|---|---|
| $B(n,p)$ 二项分布 | $p_k=C_n^k p^k(1-p)^{n-k}$ | $np$ | $np(1-p)$ | |
| $G(p)$ 几何分布 | $p_k=(1-p)^{k-1}p$ | $\frac{1}{p}$ | $\frac{1-p}{p^2}$ | 无记忆性. |
| 超几何分布 | $p_k=\frac{C_{N_1}^{k}C_{N-N_1}^{n-k}}{C_N^n}$ | $n\cdot\frac{N_1}{N}$ | $n\frac{N_1}{N}\frac{N_2}{N}(1-\frac{n-1}{N-1})$ | |
| 为什么像泊松分布？ ~~负二项分布~~ | ~~$p_k=C_{n-1}^{k-1}(1-p)^{n-k}p^k$~~ | $u=$ | $\int_0^{+\infty}x^{-\frac{1}{2}}e^{x}$ | |
| $P(\lambda)$ 泊松分布 | $p_k=\frac{\lambda^k}{k!}e^{-\lambda}$ | $\lambda$ | $\lambda$ | $\int_0^{+\infty}$ |
| 负二项分布 | $p_k=C_{k+r-1}^{k}(1-p)^kp^r$ | $\frac{r(1-p)}{p}$ / $\frac{r}{p}$ | $\frac{r(1-p)}{p^2}$ | |

常用连续分布

| 名称 | 密度函数 | 概率分布 | 期望 | 方差 |
|---|---|---|---|---|
| 指数分布 $X\sim e(\lambda)$ | $f(x)=\begin{cases}\lambda e^{-\lambda x} & x\ge0\\ 0 & x\le0\end{cases}$ | $F(x)=\begin{cases}1-e^{-\lambda x} & x>0\\ 0 & x\le0\end{cases}$ | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ |
| 正态分布 $X\sim N(u,\sigma^2)$ | $f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-u)^2}{2\sigma^2}}$ | $\phi(x)$ | $u$ | $\sigma$ |
| 均匀分布 $X\sim U(a,b)$ | $f(x)=\begin{cases}\frac{1}{b-a} & a<x<b\\ 0 & 其它\end{cases}$ | $F(x)=\begin{cases}\frac{1}{b-a}(x-a) & a<x<b\\ ? & 其它\end{cases}$ | $\frac{a+b}{2}$ | $\frac{(b-a)^2}{12}$ |

随机变量的函数

- 离散型：一一对应即可
- 连续型：$f_X(x)$ 求解 $f_Y(y)$ 先求 $F_Y(y)$. $F_Y(y)=P(g(X)\le y)$

## 样本及抽样分布

总体 ↙ ↖推断

样本 → 样本观察值

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

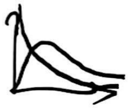
样本原点矩、中心矩

双侧、单侧分位数

$P\{|X|>T_\alpha\}=\alpha$

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

常用统计分布：① $\chi^2$分布



② F分布 $F=\frac{X/n_1}{Y/n_2}$ $\quad X\sim\chi^2(n_1)\ Y\sim\chi^2(n_2)$

X、Y独立、

③ t分布 $t=\frac{X}{\sqrt{Y/n}}$ $\quad X\sim N(0,1)\ Y\sim\chi^2(n)$

X、Y独立、

重要定理：对于正态总体 $X\sim N(\mu,\sigma^2)$：

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

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

③ $\bar{X}$与$S^2$相互独立.

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

## 参数估计

例：$E(S)=\sqrt{ES^2-DS}<\sigma$

eg：无偏估计的函数未必无偏

### 点估计

评价准则：
- 无偏、渐近无偏性
- 有效性
- 相合性 $\lim_{n\to\infty}P\{|\hat{\theta}_n-\theta|>\varepsilon\}=0$

切比雪夫不等式证明

估计方法：
- 矩估计
- 极大似然 $L(\theta)=\prod_{i=1}^{n}p(x_i;\theta)$

均值估计算可能取1/n

### 区间估计

一个总体

| 待估参数 | | 枢轴量 | 分布 | 双侧置信区间的上、下限 |
|---|---|---|---|---|
| $\mu$ | $\sigma^2$未知 | $\frac{\bar{X}-\mu}{\sqrt{S^2/n}}$ | $t(n-1)$ | $(\bar{X}-t_{\frac{\alpha}{2}}\sqrt{\sigma^2/n},\ \bar{X}+t_{\frac{\alpha}{2}}\sqrt{\sigma^2/n})$ |
| | $\sigma^2$已知 | $\frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}}$ | $N(0,1)$ | $(\bar{X}-Z_{\frac{\alpha}{2}}\sqrt{\sigma^2/n},\ \bar{X}+Z_{\frac{\alpha}{2}}\sqrt{\sigma^2/n})$ |
| $\sigma^2$ | $\mu$已知 | $\cancel{\frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}}}$ $\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2}$ | $\cancel{N(0,1)}$ $\chi^2(n)$ | $\left(\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\frac{\alpha}{2}}(n)},\ \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\frac{\alpha}{2}}(n)}\right)$ |
| | $\mu$未知 | $\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2}$ | $\chi^2(n-1)$ | $\left(\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\chi^2_{\frac{\alpha}{2}}(n-1)},\ \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right)$ |

两个总体

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

考虑：$\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}\sim F$

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

$\mu_1-\mu_2$：
- $\sigma_1^2$、$\sigma_2^2$已知 $\quad\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}\sim N(0,1)$
- 均未知 $\quad\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_w\sqrt{\frac{1}{m}+\frac{1}{n}}}\sim t(m+n-2)$

$\frac{\sigma_1^2}{\sigma_2^2}$：
- $\mu_1$、$\mu_2$已知 $\quad\frac{n\sum_{i=1}^{m}(X_i-\mu_1)^2/\sigma_1^2}{m\sum_{j=1}^{n}(Y_j-\mu_2)^2/\sigma_2^2}\sim\cancel{F(n_1-1,n_2-1)}$ $F(m,n)$
- 未知 $\quad\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}\sim F(m-1,n-1)$

## 假设检验

1° 提出原假设$H_0$及备择假设$H_1$

2° 确定显著性水平α及样本容量n

3° 确定统计检验量及拒绝域形式 {双边 / 单边

4° 按P{当$H_0$为真时拒绝$H_0$}≤α 求出拒绝域

5° 取样，根据观察值判断是否接受$H_0$

## §5 多维随机变量及其分布

### 联合分布函数

对于二维：(1) 不减

(2) $F(-\infty,y)=0$

$F(x,-\infty)=0$

$F(-\infty,-\infty)=0$

$F(+\infty,+\infty)=1$

(3) 关于 $x,y$ 右连续

(4) 正定性

### 边缘分布

$F_X(x)=F(x,+\infty)$

$F_Y(y)=F(+\infty,y)$

- 离散型二维分布　联合分布律 → 边缘分布律　独立性
- 连续型二维分布　联合概率密度 → 边缘密度函数

独立性：$f(x,y)=f_X(x)f_Y(y)$

### 常见二维随机变量分布：

①均匀分布

②二维正态分布

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

$f_X(x)=N(\mu_1,\sigma_1^2)$

$f_Y(x)=N(\mu_2,\sigma_2^2)$

$\rho=0$ 时 $X,Y$ 相互独立.

联合分布 $\underset{\nleftarrow}{\rightarrow}$ 边缘分布

联合分布 $\underset{\nleftarrow}{\rightarrow}$ 条件分布

$\left.\begin{matrix}\text{边缘分布}\\ \text{条件分布}\end{matrix}\right\}\rightarrow$ 联合分布

### 条件分布：

离散型　条件分布律

连续型　$f_{X|Y}(x|y)=\dfrac{f(x,y)}{f_Y(y)}$

$$P\{X\le x|Y=y\}=F_{X|Y}(x|y)=\int_{-\infty}^{x}\frac{f(x,y)}{f_Y(y)}dx$$

### 两个随机变量的分布(考虑卷积)

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

$Z=\frac{Y}{X}$　$\int_{-\infty}^{\infty}|x|f(x,xz)dx$

$Z=XY$　$\int_{-\infty}^{\infty}\frac{1}{|x|}f(x,\frac{z}{x})dx$

$M=\max\{X,Y\}$　$N=\min\{X,Y\}$

可加性：二项分布、泊松分布、正态分布、$\chi^2$分布

### 数字特征

期望、方差、协方差

$$cov(X,Y)=E[X-E(X)][Y-E(Y)]=E(XY)-E(X)E(Y)$$

独立 $\underset{\nleftarrow}{\rightarrow}$ 协方差等于0　分配对模型

相关系数　$corr(X,Y)=\dfrac{cov(X,Y)}{\sqrt{D(X)D(Y)}}$

~~矩阵~~ 协方差矩阵 $cov(X)$ 或 $\Sigma$ 对称、非负定.

对于二维正态分布：不相关与独立等价.

### 大数定律与中心极限定理.

大数定理：依概率收敛 $X_n\xrightarrow{P}a$

$$\lim_{n\to\infty}\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}E(X_i)\right|>\varepsilon\right)=0\quad\forall\varepsilon>0$$

不要求随机变量满足独立同分布.

中心极限定理：一般情况：期望、方差有限，$\sum_{i=1}^{n}\sigma_i^2\to\infty$

独立同分布：$\dfrac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n\sigma^2}}\sim N(0,1)$