# 1 一、随机事件

## 1.1 基本概念释义

现实生活中，一个动作或一件事情，在一定条件下，所得的结果不能预先完全确定，而只能确定是多种可能结果中的一种，称这种现象为**随机现象**。

例如，抛掷一枚硬币，其结果有可能是出现正面，也有可能是出现反面；掷骰子游戏中，出现的数字可能是1,2,3,4,5,6其中的任意一个。以上这些现象都是随机现象。

使随机现象得以实现和对它观察的全过程称为随机试验，记为$E$。随机实验满足以下三个条件:

1. 可以在相同条件下重复进行；
2. 结果有多种可能性，并且所有可能结果事先已知；
3. 作一次试验究竟哪个结果出现，事先不能确定。

接下来，我们给出如下关于样本空间，样本点，随机事件等的定义。

1. 称随机试验的所有可能结果组成的集合为**样本空间**，记为$\Omega$。
2. 试验的每一个可能结果称为**样本点**，记为$\omega$。
3. 称样本空间$\Omega$中满足一定条件的子集为**随机事件**，用大写字母$A,B,C...$表示。另外，随机事件在随机试验中可能出现也可能不出现。
4. 在试验中，称一个事件发生是指构成该事件的一个样本点出现。由于样本空间$\Omega$包含了所有的样本点，所以在每次试验中，它总是发生，因此称$\Omega$为**必然事件**。
5. 空集$\phi$不包含任何样本点，且在每次试验中总不发生，所以称为**不可能事件**。

---

以上各种概念，云里雾里的，下面举个栗子就清楚了。

掷骰子游戏中，我们知道出现的结果可能是1,2,3,4,5,6其中的任意一个数字。那么出现任何一个数字，都可以成为一个样本点；随机事件是什么呢，就是一些样本点的的集合，当然了，是在一定条件下。比如，出现的数字是偶数的结果。那么2,4,6就够成了一个随机事件$A=\{2,4,6\}$。样本空间就是1到6的六个数字$\Omega=\{1,2,3,4,5,6\}$。可以看到$A$是$\Omega$的一个子集。空集可以定义$\phi$为结果的数字大于6，显然是不可能出现的。

---

## 1.2 概率

### 1.2.1 定义：

随机试验$E$的样本空间为$\Omega$，对于每个事件$A$，定义一个实数$P(A)$与之对应，若函数$P(.)$满足条件：

1. 对每个事件$A$，均有$0<P(A)<=1$;

2. $P(\Omega)=1$;
3. 若事件$A_1,A_2,A_3,\ldots$两两互斥，即对于$i,\ j=1,2,\ldots,\ i\neq j, A_i\cap A_j=\phi$，均有
   $P(A_1\cup A_2\cup\ldots)=P(A_1)+P(A_2)+\ldots$

则称$P(A)$为事件$A$的概率。

### 1.2.2 主要性质：

1. 对于任一事件$A$，均有$P(\overline{A})=1-P(A)$.
2. 对于两个事件$A$和$B$，若$A\subset B$，则有

$$P(B-A)=P(B)-P(A), P(B)>P(A).$$

1. 对于任意两个事件$A$和$B$，有

$$P(A\cup B)=P(A)+P(B)-P(A\cap B).$$

我们仍然以掷骰子游戏举例。

掷骰子中，1,2,3,4,5,6出现的概率均为1/6。 我们令 $A=\{1,2\}, B=\{1,2,3\}$。那么有$\overline{A}=\{3,4,5,6\}$。可以看到，出现1或2的概率为1/3，即$P(A)=1/3$；出现1或2或3的概率为1/2，即$P(B)=1/2$。根据性质我们有

1. $P(\overline{A})=1-P(A)=1-1/3=2/3$,也就是出现3或4或5或6的概率；
2. $P(B-A)=P(B)-P(A)=1/2-1/3=1/6$,也就是出现3的概率；
3. $P(A\cup B)=P(A)+P(B)-P(A\cap B)=1/3+1/2-1/3=1/2$,也就是出现的1或2或3，也就是事件$B$的概率；因为$A\subset B$。这里的$A\cap B=A=\{1,2\}$。

## 1.3 古典概型

我们将掷骰子游戏进行推广，设随机事件$E$的样本空间中只有有限个样本点，即$\Omega=\{\omega_1,\omega_2,\ldots,\omega_n\}$，其中，$n$为样本点的总数。每个样本点$\omega_i(i=1,2,\ldots,n)$出现是等可能的，并且每次试验有且仅有一个样本点发生，则称这类现象为古典概型。若事件$A$包含个$m$个样本点，则事件$A$的概率定义为：

$P(A)=\frac{m}{n}=\frac{\text{事件}A\text{包含的基本事件数}}{\text{基本事件总数}}$。

古典概型是不是很简单，接下来我们基于古典概型进行例题的推广。坐好了，下面的知识点会涉及排列组合。

假设有$k$个不同颜色的球，每个球以同样的概率$1/l$落到$l$个格子$(l>=k)$的每个中，且每个格子可容纳任意多个球。问，分别求出如下两个事件$A$和$B$的概率。

1. $A$:指定的$k$个格子中各有一个球；
2. $B$:存在$k$个格子，其中各有一个球。

我们思考一下，由于每个球可以平均地落入$l$个格子中的任一个，并且每一个格子中可落入任意多个球，所以$k$个球落入$l$个格子中的分布情况相当于从$l$个格子中选取$k$个的可重复排列，故样本空间共有$l^k$种等可能的基本结果。

所以，事件$A$所含基本结果数应是$k$个球在指定的$l$个格子中的全排列数，即$k!$那么有

$$P(A)=\frac{k!}{l^k}$$

为了算出事件 $B$ 所含的基本事件数，我们可以分两步进行：因为 $l$ 个格子可以是任意选取的，故可先从 $l$ 个格子中任意选出 $k$ 个出来，那么选法共有 $C_l^k$ 种。对于每种选定的 $k$ 个格子，依上述各有一个球的推理，则有 $k!$ 个基本结果，故B含有 $C_l^k * k!$ 个基本结果。那么有

$$P(B)=\frac{C_l^k*k!}{l^k}=\frac{l!}{l^k*(l-k)!}$$

我们把上述例子应有到具体的问题中，概率论的历史上有一个颇为著名的问题**生日问题**：求 $k$ 个同班同学没有两人生日相同的概率。

如果把这 $k$ 个同学看作上例中的 $k$ 个球，而把一年365天看作格子，即 $l=365$ ，则上述的 $P(B)$ 就是所要求的概率。我们令 $k=40$ 时，利用上面的公式，则 $P(B)=0.109$ 。换句话说，40个同学中至少两个人同一天过生日的概率是：$P(\overline{B})=1-0.109=0.891$ 。其概率大的出乎意料。

这讲内容更多地是对概念知识的理解，不太涉及软件的实现，给出简单的 $P(B)$ Python实现：

```python
#我们采用函数的递归的方法计算阶乘：
def factorial(n):
    if n == 0:
        return 1;
    else:
        return (n*factorial(n-1))

l_fac = factorial(365);          #l的阶乘
l_k_fac = factorial(365-40)      #l-k的阶乘
l_k_exp = 365**40                #l的k次方

P_B =  l_fac /(l_k_fac * l_k_exp)     #P(B)
print("事件B的概率为：",P_B)
print("40个同学中至少两个人同一天过生日的概率是：",1 - P_B)

```

## 1.4 条件概率

研究随机事件之间的关系时，在已知某些事件发生的条件下考虑另一些事件发生的概率规律有无变化及如何变化，是十分重要的。我们先给出定义，然后进行例子的讲解与描述。

1. 定义：

设 $A$ 和 $B$ 是两个事件，且 $P(B)>0$ ，称 $P(A|B)=\frac{P(AB)}{P(B)}$ 为在事件 $B$ 发生的条件下，事件 $A$ 发生的概率。

1. 例子：

某集体中有 $N$ 个男人和 $M$ 个女人，其中患色盲者男性 $n$ 人，女性 $m$ 人。我们用 $\Omega$ 表示该集体， $A$ 表示其中全体女性的集合， $B$ 表示其中全体色盲者的集合。如果从 $\Omega$ 中随意抽取一人，则这个人分别是女性、色盲者和同时既为女性又是色盲者的概率分别为：

$P(A)=\frac{m}{m+N}$ $P(B)=\frac{m+n}{m+N}$ $P(A,B)=\frac{n}{m+N}$

$P(A)=\frac{M}{M+N}, P(B)=\frac{m+n}{M+N}, P(AB)=\frac{m}{M+N}$

$P(B|A)=\frac{P(AB)}{P(A)}=\frac{m}{M}$

如果限定只从女性中随机抽取一人**(即事件 $A$ 已发生)**，那么这个女人为色盲者的**(条件)**概率为

$P(B|A)=\frac{m}{M}=\frac{P(AB)}{P(A)}$

## 1.5 全概率公式和贝叶斯公式

1. **准备知识：**首先我们看一下概率乘法公式和样本空间划分的定义；
   a. 由条件概率公式，可以得到**概率的乘法公式**：
   $P(AB)=P(B|A)P(A)=P(A|B)P(B)$
   b. 如果事件组，满足
      1. $B_1,B_2,\dots$ 两两互斥，即 $B_i\cap B_j=\phi,\ i\neq j,i,j=1,2,\dots$，且 $P(B_i)>0,i=1,2,\dots$
      2. $B_1\cup B_2\cup\ldots=\Omega$

   则称事件组 $B_1,B_2,\dots$ 是样本空间 $\Omega$ 的一个划分。

1. **全概率公式**
   设 $B_1,B_2,\dots$ 是样本空间 $\Omega$ 的一个划分，$A$ 为任一事件，则
   $P(A)=\sum_{i=1}^{\infty}P(B_i)P(A|B_i)$
   称为全概率公式。
   根据全概率公式和概率乘法公式，我们可以得到：
2. **贝叶斯公式**
   设 $B_1,B_2,\dots$ 是样本空间 $\Omega$ 的一个划分，则对任一事件 $A(P(A)>0)$，有
   $P(B_i|A)=\frac{P(B_iA)}{P(A)}=\frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{\infty}P(B_j)P(A|B_j)},i=1,2,\dots$
   称上式为贝叶斯公式，称 $P(B_i)(i=1,2,\dots)$ 为先验概率，$P(B_i|A)\ (i=1,2,\dots)$ 为后验概率。

有点懵……，不急，我们看下面的例子吧。

在实际中，常取对样本空间 $\Omega$ 的有限划分 $B_1,B_2,\dots,B_n$。$B_i$ 视为导致试验结果 $A$ 发生的“原因”，而 $P(B_i)$ 表示各种“原因”发生的可能性大小，故称为先验概率；$P(B_i|A)$ 则反应当试验产生了结果 $A$ 之后，再对各种“原因”概率的新认识，故称为后验概率。

假定用血清甲胎蛋白法诊断肝癌。用 $C$ 表示被检验者有肝癌这一事件，用 $A$ 表示被检验者为阳性反应这一事件。当前有肝癌的患者被检测呈阳性反应的概率为0.95。即 $P(A|C)=0.95$。当前非肝癌的患者被检测呈阴性反应的概率为0.9。即 $P(\overline{A}|\overline{C})=0.90$。若某人群中肝癌患者概率为0.0004，即 $P(C)=0.0004$，现在有一人呈阳性反应，求此人确为肝癌患者的概率是多少？

**解**：

$P(C|A)=\frac{P(C)P(A|C)}{P(C)P(A|C)+P(\overline{C})P(A|\overline{C})}=\frac{0.0004*0.95}{0.0004*0.95+0.9996*0.1}=0.0038$

**贝叶斯公式也是在机器学习中朴素贝叶斯的核心，请大家予以重视~!**

# 2 二、随机变量

## 2.1 随机变量及其分布

1. 随机变量定义：

设 $E$ 是随机试验，$\Omega$ 是样本空间，如果对于每一个 $\omega \in \Omega$ 。都有一个确定的实数 $X(\omega)$ 与之对应，若对于任意实 $x \in R$，有 $\{\omega: X(\omega) < x\} \in F$ ，则称 $\Omega$ 上的单值实函数 $X(\omega)$ 为一个随机变量。

从定义可知随机变量是定义在样本空间 $\Omega$ 上，取值在实数域上的函数。由于它的自变量是随机试验的结果，而随机试验结果的出现具有随机性，因此，随机变量的取值也具有一定的随机性。这是随机变量与普通函数的不同之处。

描述一个随机变量，不仅要说明它能够取那些值，而且还要关心它取这些值的概率。因此，接下来引入随机变量的分布函数的概念。

1. 随机变量的分布函数定义：

设 $X$ 是一个随机变量，对任意的实数 $x$ ，令

$$F(x) = P\{X <= x\}, x \in (-\infty, +\infty)$$

则称 $F(x)$ 为随机变量 $x$ 的分布函数，也称为概率累积函数。

直观上看，分布函数 $F(x)$ 是一个定义在 $(-\infty, +\infty)$ 上的实值函数， $F(x)$ 在点 $x$ 处取值为随机变量 $X$ 落在区间 $(-\infty, +x]$ 上的概率 。分布函数（概率累积函数）很好理解，就是在一个区间范围内概率函数的累加。这个区间就是负无穷到当前节点。

## 2.2 离散型随机变量

如果随机变量 $X$ 的全部可能取值只有有限多个或可列无穷多个，则称 $X$ 为离散型随机变量。掷骰子的结果就是离散型随机变量。

对于离散型随机变量 $X$ 可能取值为 $x_k$ 的概率为：

$$P\{X = x_k\} = p_k, k = 1, 2, \dots$$

则称上式为离散型随机变量 $X$ 的分布律。

我们可以用下表来表示分布律：

| $X$ | $x_1$ | $x_2$ | ..... | $x_n$ | ..... |
|---|---|---|---|---|---|

| $p_k$ | $p_1$ | $p_2$ | ..... | $p_n$ | ..... |
|---|---|---|---|---|---|

离散型随机变量的分布函数为：

$$F(x)=P\{X<=x\}=\sum_{x_k<=x}P\{X=x_k\}=\sum_{x_k<=x}P_k$$

## 2.3 常见的离散型分布

### 2.3.1 伯努利实验，二项分布

1. 定义：

   如果一个随机试验只有两种可能的结果 $A$ 和 $\overline{A}$，并且

$$P(A)=p,\ P(\overline{A})=1-p=q$$

其中，$0<p<1$，则称此试验为Bernoulli(伯努利)试验. Bernoulli试验独立重复进行 $n$ 次，称为 $n$ 重伯努利试验。

看例子

从一批产品中检验次品，在其中进行有放回抽样 $n$ 次，抽到次品称为“成功”，抽到正品称为“失败“，这就是 $n$ 重Bernoulli试验。

设

$$A=\{n重伯努利试验中A出现k次\}$$

则

$$P(A_k)=C_n^kp^k(1-p)^{n-k},k=0,1,2,\dots n.$$

这就是著名的二项分布，常记作 $B(n,\ k)$。

解释：一共抽了 $n$ 次，$k(k<n)$ 次抽中了 $A$,概率为 $p$,那么 $n-k$ 次抽中了非 $A$，概率为 $1-p$ 组合的次数就是 $C_n^k$。所以 $P(A_k)=C_n^kp^k(1-p)^{n-k},k=0,1,2,\dots n.$

1. 分布函数：

若随机变量 $X$ 的分布律为：

$$P\{X=k\}=C_n^kp^k(1-p)^{n-k},k=0,1,2,\dots n.$$

其分布函数为：

$$F(x)=\sum_{k=}^{[x]}C_n^kp^k(1-p)^{n-k},k=0,1,2,\dots n.$$

其中，$[x]$ 表示下取整，即不超过 $x$ 的最大整数。

## 2.4 随机变量的数字特征

### 2.4.1 数学期望

1. 离散型：设离散型随机变量 $X$ 的分布律为 $P\{X=x_i\}=p_i, i=1,2,\ldots,$ 若级数 $\sum_i |x_i|p_i$ 收敛，（收敛指会聚于一点，向某一值靠近，相对于发散）。则称级数 $\sum_i x_i p_i$ 的和为随机变量 $X$ 的数学期望。记为 $E(X)$，即：

$$E(X)=\sum_i x_i p_i$$

1. 设连续型随机变量 $X$ 的概率密度函数为 $f(x)$，若积分 $\int_{-\infty}^{+\infty}|x|f(x)\,dx$ 收敛，称积分 $\int_{-\infty}^{+\infty}xf(x)\,dx$ 的值为随机变量 $X$ 的数学期望，记为 $E(X)$，即：

$$E(X)=\int_{-\infty}^{+\infty}xf(x)\,dx$$

```
1  $E(X)$ 又称为均值。
```

数学期望代表了随机变量取值的平均值，是一个重要的数字特征。数学期望具有如下性质：

1. 若 $c$ 是常数，则 $E(c)=c$；
2. $E(aX+bY)=aE(X)+bE(Y)$，其中a, b为任意常数；
3. 若 $X,Y$ 相互独立，则 $E(XY)=E(X)E(Y)$；（相互独立就是没有关系，不相互影响）。

###

1. 设 $X$ 为随机变量，如果 $E\{[X-E(X)]^2\}$ 存在，则称 $E\{[X-E(X)]^2\}$ 为 $X$ 的方差。记为 $Var(X)$，即：

$$Var(X)=E\{[X-E(X)]^2\}$$

并且称 $\sqrt{Var(X)}$ 为 $X$ 的标准差或均方差。

方差是用来描述随机变量取值相对于均值的离散程度的一个量，也是非常重要的数字特征。方差有如下性质：

1. 若 $c$ 是常数，则 $Var(c)=0$；
2. $Var(aX+b)=a^2Var(X)$，其中a, b为任意常数；
3. 若 $X,Y$ 相互独立，则 $Var(X+Y)=Var(X)+Var(Y)$。

$$Var[aX+b]=E[(aX+b-E[aX+b])^2]$$
$$=E[(aX+b-aE[X]-b)^2]$$
$$=E[(aX-aE[X])^2]$$
$$=E[a^2(X-E[X])^2]$$
$$=a^2E[(X-E[X])^2]=a^2Var[X]$$

### 2.4.2 方差

协方差和相关系数都是描述随机变量 $X$ 与随机变量 $Y$ 之间的线性联系程度的数字量。

1. 设 $X,Y$ 为两个随机变量，称 $E\{[X-E(X)][Y-E(Y)]\}$ 为 $X$ 和 $Y$ 的协方差，记为 $Cov(X,Y)$，即：

证明：$Cov(X,Y)=E[XY]-E[X]\cdot E[Y]$

$Cov[X,Y]=E[(X-\mu)(Y-\nu)]=E[XY-X\nu-Y\mu+X\nu]$

$=E[XY]-E[X\mu]-E[Y\nu]+E[\mu\nu]$

$=E[XY]-\mu\nu=E[XY]-E[X]\cdot E[Y]$ → 类比 $Var[X]=E[X^2]-E[X]^2$

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

协方差有如下性质：

$Cov[X_1+X_2,Y]=E[(X_1+X_2-E[X_1+X_2])\cdot(Y-E[Y])]$

$=E[((X_1-E[X_1])+(X_2-E[X_2]))\cdot(Y-E[Y])]$ ……

1. $Cov(X,Y)=Cov(Y,X)$；
2. $Cov(aX+b,\ cY+d)=acCov(X,\ Y)$，其中，$a,b,c,d$ 为任意常数；
3. $Cov(X_1+X_2,\ Y)=Cov(X_1,\ Y)+Cov(X_2,\ Y)$；
4. $Cov(X,\ Y)=E(X,\ Y)-E(X)E(Y)$；当 $X,Y$ 相互独立时，有 $Cov(X,\ Y)=0$；
5. $|Cov(X,\ Y)|\leq\sqrt{Var(X)}\sqrt{Var(Y)}$；
6. $Cov(X,\ X)=Var(X)$；

Cov 的性质可由定义，E，V 的性质推导

① ② ③

2. 当 $\sqrt{Var(X)}>0,\ \sqrt{Var(Y)}>0$ 时，称

相关系数即标准化后的协方差

$$\rho_{(X,Y)}=\frac{Cov(X,\ Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$$

为 $X,Y$ 的相关系数，它是无纲量的量（也就是说没有单位，只是个代数值）。

3. 基本上我们都会用相关系数来衡量两个变量之间的相关程度。相关系数在-1到1之间，小于零表示负相关，大于零表示正相关。绝对值 $|\rho(X,Y)|$ 表示相关度的大小。越接近1，相关度越大。

证明 $\rho(X,Y)\in[-1,1]$

1. 先证明 $[E(XY)]^2\leq E(X^2)E(Y^2)$

构造 $g(t)=E[(X+tY)^2]=E[X^2+t^2Y^2+2tXY]=E[X^2]+2tE[XY]+t^2E[Y^2]\geq 0$

（$c=E[X^2]$，$b=2E[XY]$，$a=E[Y^2]$）

$\therefore\Delta=b^2-4ac=4E[XY]^2-4E[X^2]E[Y^2]\leq 0$

即 $E[XY]^2\leq E[X^2]\cdot E[Y^2]$

2. 由可得 $[E[(X-E[X])\cdot(Y-E[Y])]]^2\leq E[(X-E[X])^2]\cdot E[(Y-E[Y])^2]$

$\therefore\ (Cov[X,Y])^2\leq Var[X]\cdot Var[Y]$

$\therefore\ \left(\frac{Cov[X,Y]}{\sqrt{Var[X]\cdot Var[Y]}}\right)^2\leq 1$

$\therefore\ -1\leq\frac{Cov[X,Y]}{\sqrt{Var[X]\cdot Var[Y]}}\leq 1$

$-\sqrt{Var[X]}\cdot\sqrt{Var[Y]}\leq Cov[X,Y]\leq\sqrt{Var[X]}\cdot\sqrt{Var[Y]}$

$\therefore\ |Cov[X,Y]|\leq\sqrt{Var[X]}\cdot\sqrt{Var[Y]}$