我们做**随机实验(ramdom trial)**，记为 E ，把实验的所有结果的集合叫作**样本空间(sample sapce)**，记为 Ω 。

随机事件(random event)是样本空间的子集，简单称为**事件。**

频率 P

那么，我们假设在同一条件下进行了 n 次实验，再假设随机事件 A 在实验中发生了 k 次，那么就事件的**频率**为：

$$f_n(A)=\frac{k}{n}$$

当 n 很大的时候，频率 $\frac{k}{n}$ 趋于某一数值 p ，则称 p 为事件 A 发生的**概率**，记为：

$$P(A)=p$$

公理化的定义是说，$P(A)$ 满足以下公理：

- 非负性：$P(A)\geq 0$
- 规范性：$P(\Omega)=1$
- 可数可加性：$P(\cup_{n=1}^{\infty}A_n)=\sum_{n=1}^{\infty}P(A_n)$

条件概率 $P(A|B)=\frac{P(AB)}{P(B)}$

我们说，在事件 B 发生的前提下，事件 A 发生的**条件概率**为:

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

条件概率 $P(A|B)$ 满足以下公理：

- 对任一事件 A ，有 $P(A|B)\geq 0$
- $P(\Omega|B)=1$
- $P(\cup_{i=1}^{\infty}A_i|B)=\sum_{i=1}^{\infty}P(A_i|B)$

另外，**乘法公式**是说，当 $P(B)>0$ 时，则有：

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

全概率公式 $P(B)=\sum_{i=1}^{n}P(A_i)P(B|A_i)$

我们定义 $A_1,A_2,...,A_n$ 为样本空间 Ω 的一个**划分**，它满足：

- 划分中的任两个事件之间不相容
- 划分中的所有事件的总和构成样本空间

那么，这里我们认定事件 B 为样本空间 Ω 中的任意事件，因为 $P(B\Omega)=P(B)*P(\Omega)=P(B)*1=P(B)$

，所以这里给出公式推理：

$$
\begin{aligned}
P(B) = P(B\Omega) &= P(B(A_1 \cup A_2 \cup \cdots \cup A_n)) \\
&= P(BA_1 \cup BA_2 \cup \cdots \cup BA_n) \\
&= P(BA_1) + P(BA_2) + \cdots + P(BA_n) \\
&= P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n)
\end{aligned}
$$

贝叶斯公式 $P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_i)P(A_j)}$

贝叶斯公式由条件概率和全概率公式组合而来，推理如下：

$$
\begin{aligned}
P(A_i|B) &= \frac{P(A_iB)}{P(B)} = \frac{P(BA_i)}{P(B)} \\
&= \frac{P(B|A_i)P(A_i)}{P(B)} \\
&= \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)}
\end{aligned}
$$

一般来说，我们将划分 $A_1, A_2, \ldots, A_n$ 作为已知的结果，是说我们通过实验或者以往的信息经验之类的得到了 $P(A_j)$ 的值，所以我们称 $P(A_j)$ 为**先验概率**。此外，我们称 $P(A_i|B)$ 为**后验概率**，因为 $P(A_i|B)$ 是说在事件 B 发生后，A_i 再发生的概率。

通俗来说，就是我们通过了那么多的事件 $A_1, A_2, \ldots, A_n$ 得到了结果，也就是事件 B 的概率，但是我们还想要知道这个结果，也就是事件 B 发生的情况下，某个 A_j 发生的概率是多少这样。

分布函数 $F(x) = P\{X \le x\}$

我们这里给出一个函数为：

$$X = X(A)$$

这里，我们把样本空间 Ω 中的每一个结果，或者说每一个事件都放入函数里，得到一个实数，比如 $X_1 = X(A_1), X_2 = X(A_2), \ldots, X_n = X(A_n)$ 。

这样做的好处是，我们将一些实验的结果用数字进行替代，比如，我们要在一个装有红、绿、蓝小球的箱子里摸球，我们可以用数字来替代红、绿、蓝的结果，$X(\text{红}) = 1, X(\text{绿}) = 2$，$X(\text{蓝}) = 3$。

因为我们的函数 X 的值会随着实验的不同结果而变化，所以我们称 X 函数为**随机变量(random variable)。**

当然，我们一般用区间对 X 的值进行描述，因为有时候不能将每一个值都列出来，所以我们会说随机变量 X 的取值落在区间 $(x_1, x_2]$ 的概率，就是要求 $P\{x_1 \le X \le x_2\}$ 的值。

求 $P\{x_1 \le X \le x_2\}$ 的值就相当于要计算 $P\{X \le x_2\} - P\{X \le x_1\}$ 的值，那么就很容易知道我们其实是要研究 $P\{X \le x\}$ 的概率问题了。因为它的值也是随着不同的 x 而变化的，所以我们叫 $P\{X \le x\}$ 为$P\{X \le x\}$ ，这里给出它的公式：

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

分布函数有以下特点：

- $F(x_2)-F(x_1)=P\{x_1<X\leq x_2\}>0$
- $0\leq F(x)\leq 1$

我们在这里对随机变量有个区分：

- 离散型随机变量：随机变量的取值为有限个或者可数无穷多个
- 连续型随机变量：随机变量的取值连续地充满某个区间

离散型随机变量-两点分布 $X\sim(0-1)$

当随机变量 X 的取值只有 x_1 和 x_2 这两个结果时，它的分布为：

$$\begin{aligned}&P\{X=x_1\}=p\quad,\\&P\{X=x_2\}=1-p\quad,\quad 0<p<1\end{aligned}$$

我们称 X 服从参数为 p 的**两点分布**，也叫 (0-1) 分布，记作 $X\sim(0-1)$。

离散型随机变量-二项分布 $X\sim b(n,p)$

当随机变量 X 的分布满足：

$$P\{X=k\}=C_n^k p^k(1-p)^{n-k}$$

则称 X 为服从参数为 n , p 的**二项分布(binomial distribution)**，记作 $X\sim b(n,p)$。

一般我们会使用 **泊松(Posisson)定理** 来进行近似计算，这里做个简单介绍。

设 $np_n=\lambda$ ，对任意非负整数有：

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

离散型随机变量-泊松分布 $X\sim P(\lambda)$

当随机变量 X 的分布满足：

$$P\{X=k\}=\frac{\lambda^k e^{-\lambda}}{k!}\quad,\quad k=0,1,2,\ldots,$$

则称 X 为服从参数为 λ 的泊松分布(poisson distribution)，记作 $X\sim P(\lambda)$，其中 λ 为常数。

概率密度函数 $f(x)$

这里介绍一个概念，**概率密度函数(density sunction)。**它可以用来描述随机变量 X 的分布函数 $F(x)$：

$$F(x)=\int_{\infty}^{x}f(t)dx$$

它有以下特点:

- $f(x) \geq 0$
- $\int_{-\infty}^{+\infty} f(x)dx = 1$
- $P\{x_1 < X < x_2\} = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(t)dx$

连续型随机变量-均匀分布 $X \sim U(a,b)$

当随机变量 X 具有概率密度:

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

则称 X 在区间 (a,b) 上服从**均匀分布(unniform distribution)**，记作 $X \sim U(a,b)$

积分求得 X 的分布函数:

$$F(x) = \begin{cases} 0, & x < a, \\ \dfrac{x-a}{b-a}, & a \leq x < b, \\ 1, & x \geq b. \end{cases}$$

连续型随机变量-指数分布 $X \sim E(\lambda)$

当随机变量 X 具有概率密度:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

则称 X 服从 λ 的**指数分布(exponential distribution)**，记作 $X \sim E(\lambda)$，其中 λ 为常数。

积分求得 X 的分布函数:

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

连续型随机变量-正态分布 $X \sim N(\mu, \sigma^2)$

当随机变量 X 具有概率密度:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < +\infty$$

则称 X 服从参数为 μ , σ 的**正态分布(normal distribution)**，记作 $X \sim N(\mu, \sigma^2)$， 其中 μ 和 $\sigma(\sigma > 0)$ 为常数。

积分求得 X 的分布函数:

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

特别的，当 $\mu=0,\sigma=1$时，我们称 X 服从标准正态分布 $N(0,1)$，这时它的概率密度表示为：

$$\varphi(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{x^2}{2}}$$

分布函数表示为：

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

一般地，若 $X\sim N(\mu,\sigma^2)$，那么 $\frac{X-\mu}{\sigma}\sim N(0,1)$，我们可以通过正态函数表来计算正态分布：

$$\begin{aligned}
P\{\mu-\sigma<X<\mu+\sigma\}=\phi(1)-\phi(-1)=2\phi(1)-1=0.6826\\
P\{\mu-2\sigma<X<\mu+2\sigma\}=\phi(2)-\phi(-2)=2\phi(2)-1=0.9544\\
P\{\mu-3\sigma<X<\mu+3\sigma\}=\phi(3)-\phi(-3)=2\phi(3)-1=0.9974
\end{aligned}$$

联合分布函数 $F(x,y)=P\{X\leq x,Y\leq y\}$

在这里，我们考虑两个随机变量 $X(e)$ 和 $Y(e)$ 的组合。我们把 $(X(e),Y(e))$ 称为**二维随机向量(2-dimensional random vector)**，简单记作 (X,Y)。

显然，我们可以得到二维随机向量 (X,Y) 的分布函数，或者说，随机变量 X 和随机变量 Y 的**联合分布函数**：

$$F(x,y)=P\{X\leq x,Y\leq y\}$$

边缘分布函数 $F_X(x),F_Y(y)$

我们说到联合分布函数是二维随机变量 (X,Y) 的分布函数，自然随机变量 X 和 Y 是有分布函数的，那么我们通过联合分布函数来求得变量 X 和 Y 的分布函数，就可以得二维随机变量 (X,Y) 关于 X 和 Y 的边缘分布函数(marginal distribution function)：

$$\begin{aligned}
F_X(x)=P\{X\leq x\}=P\{X\leq x,Y<+\infty\}=F(x,+\infty)\\
F_Y(y)=P\{Y\leq y\}=P\{X<+\infty,Y\leq y\}=F(+\infty,y)
\end{aligned}$$

数学期望 $E(X)$

我们现在知道一个离散型变量 X 的分布律为：

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

如果满足条件 $\sum_{k=1}^{\infty}x_kp_k$ 绝对收敛，那么有**数学期望(mathematical expectation)**，记作 $E(X)$，即：

$$E(X)=\sum_{k=1}^{\infty}x_kp_k$$

当然，如果是一个连续型的随机变量 X 的话，我们就假设它有概率密度函数 $f(x)$。

如果满足积分 $\int_{-\infty}^{+\infty}xf(x)dx$ 绝对收敛，那么同样有数学期望为：

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

数学期望有一些性质：

- $E(c)=c$，其中 c 为常数；
- $E(cX)=cE(X)$；
- $E(X+Y)=E(X)+E(Y)$；
- $E(XY)=E(X)E(Y)$，其中 X,Y 相互独立。

	$(0-1)$**分布**	**二项分布**	**泊松分布**	**均匀分布**	**指数分布**	**正态分布**
$E(X)$	p	np	λ	$\frac{a+b}{2}$	$\frac{1}{\lambda}$	μ

方差 $D(X)$

数学期望描述了随机变量取值的“平均数”，而**方差(variance)**是用来度量随机变量取值的分散程度的，记作 $D(X)$，即：

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

其中，我们称 $\sqrt{D(X)}$ 为随机变量 X 的**标准差(standard deviation)**，或**均方差(mean square deviation)**，记作 $\sigma(X)$。

方差有一些性质：

- $D(c)=0$，其中 c 为常数；
- $D(cX)=c^2D(X)$；
- $D(X\pm Y)=D(X)+D(Y)\pm 2E[(X-E(X))(Y-E(Y))]$；
- $D(X\pm Y)=D(X)+D(Y)$，其中 X,Y 相互独立。

	$(0-1)$**分布**	**二项分布**	**泊松分布**	**均匀分布**	**指数分布**	**正态分布**
$D(X)$	$p(1-p)$	$np(1-p)$	λ	$\frac{(a+b)^2}{12}$	$\frac{1}{\lambda^2}$	σ^2

协方差 $cov(X,Y)$

数学期望和方差反映的都是随机变量自身的内容，这里我们考虑随机变量相互之间的影响，一般会使用**协方差(convariance)** 来描述，即：

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

其中，我们称 $\frac{cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$ 为随机变量 X,Y 的**相关系数(correlation corfficient)**，或标准协方差(standard convariance)，记作 ρ_{XY}，即：

$$\rho_{XY} = \frac{cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

一些实用的计算公式：

$$D(X \pm Y) = D(X) + D(Y) \pm 2cov(X,Y)$$
$$cov(X,Y) = E(XY) - E(X)E(Y)$$