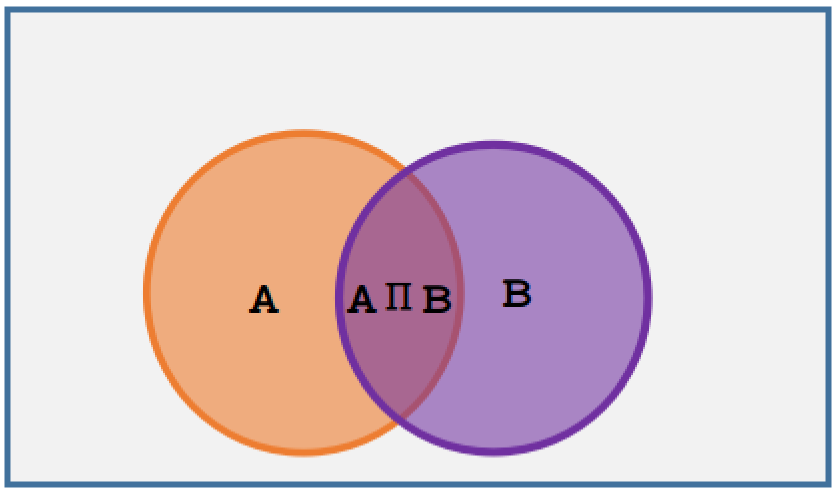
条件概率、全概率公式和贝叶斯公式

一、条件概率公式

举个例子，比如让你背对着一个人，让你猜猜背后这个人是女孩的概率是多少？直接猜测，肯定是只有５０％的概率，假如现在告诉你背后这个人是个长头发，那么女的概率就变为９０％。所以条件概率的意义就是，当给定条件发生变化后，会导致事件发生的可能性发生变化。

条件概率由文氏图出发，比较容易理解：



$p(A|B)$表示B发生后A发生的概率，由上图可以看出B发生后，A再发生的概率就是$\frac{P(A\cap B)}{P(B)}$，因此：

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

由：

$$p(A|B)=\frac{P(A\cap B)}{P(B)}\Rightarrow p(A\cap B)=p(A|B)\times p(B)$$

$$\Rightarrow p(A\cap B)=P(B|A)\times P(A)$$

得：

$$p(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{p(B|A)\times P(A)}{p(B)}$$

这就是条件概率公式。

假如事件A与B相互独立，那么：

$$p(A\cap B)=P(A)\times P(B)$$

注：

相互独立：表示两个事件发生互不影响。而互斥：表示两个事件不能同时发生，（两个事件肯定没有交集）。互斥事件一定不独立（因为一件事的发生导致了另一件事不能发生）；独立事件一定不互斥，（如果独立事件互斥， 那么根据互斥事件一定不独立，那么就矛盾了），但是在概率形式上具有一些巧合性，一般地：

$$\begin{cases} p(AB)=P(A)\times P(B) & independence \\ p(AB)=0 & mutex \end{cases}$$

但是，对于两个独立事件，$p(AB)$依然可以等于0，因为事件A或者事件B发生的概率可能为0. 所以$p(AB)=0$，并不是一定表示互斥。互斥和独立的理解还是要究其真正意义，而不是表达形式。

二、全概率公式

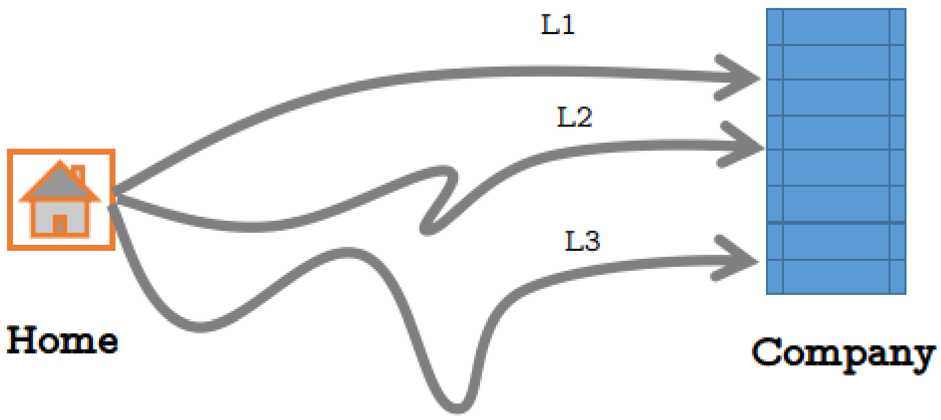
先举个例子，小张从家到公司上班总共有三条路可以直达（如下图），但是每条路每天拥堵的可能性不太一样，由于路的远近不同，选择每条路的概率如下：

$p(L_1)=0.5, p(L_2)=0.3, p(L_3)=0.2$

每天上述三条路不拥堵的概率分别为：

$p(C_1)=0.2, p(C_2)=0.4, p(C_3)=0.7$

假设遇到拥堵会迟到，那么小张从Home到Company不迟到的概率是多少？



其实不迟到就是对应着不拥堵，设事件C为到公司不迟到，事件L_i为选择第i条路，则：

$p(C)=p(L_1)\times p(C|L_1)+p(L_2)\times p(C|L_{})+p(L_3)\times p(C|L_3)$

$p(C)=p(L_1)\times p(C_1)+p(L_2)\times p(C_2)+p(L_3)\times p(C_3)$

$p(C)=0.5\times 0.2+0.3\times 0.4+0.2\times 0.7=0.36$

全概率就是表示达到某个目的，有多种方式（或者造成某种结果，有多种原因），问达到目的的概率是多少（造成这种结果的概率是多少）？

全概率公式：

设事件$L_1, L_2\cdots\cdots$是一个完备事件组，则对于任意一个事件C，若有如下公式成立：

$$p(C)=p(L_1)p(C|L_1)\cdots\cdots p(L_n)p(C|L_n)=\sum_{i=1}^{n}p(L_i)p(C|L_i)$$

那么就称这个公式为全概率公式。

三、贝叶斯公式

仍旧借用上述的例子，但是问题发生了改变，问题修改为：到达公司未迟到选择第 1 条路的概率是多少？

可不是$p(L_1)=0.5$，因为０．５这个概率表示的是，选择第一条路的时候并没有靠考虑是不是迟到，只是因为距离公司近才知道选择它的概率，而现在我们是知道未迟到这个结果，是在这个基础上问你选择第一条路的概率，所以并不是直接就可以得出的。

故有：

$$p(L_1|C)=\frac{p(C|L_1)\times p(L_1)}{p(C)}$$

$$p(L_1|C)=\frac{p(C|L_1)\times p(L_1)}{P(L_1)\times p(C|L_1)+P(L_2)\times p(C|L_2)+P(L_3)\times p(C|L_3)}$$

$$p(L_1|C)=\frac{0.2\times 0.5}{0.2\times 0.5+0.3\times 0.4+0.2\times 0.7}=0.28$$

所以选择第一条路的概率为０．２８．

贝叶斯公式就是当已知结果，问导致这个结果的第i原因的可能性是多少？执果索因！

贝叶斯公式：

在已知条件概率和全概率的基础上，贝叶斯公式是很容易计算的：

$$p(L_k|C)=\frac{p(C|L_k)\times p(L_k)}{p(C)}\quad\Rightarrow\quad p(L_k|C)=\frac{p(C|L_k)\times p(L_k)}{\sum_{i=1}^{n}p(L_i)\times p(C|L_i)}$$