数据科学基础

Foundations of Data Science

2.2 概率的计算

陈振宇
南京大学智能软件工程实验室
www.iselab.cn

条件概率

已知某一事件已发生，求另一事件发生的概率。

- 例如考虑有两个孩子的家庭，假定男女出生率一样，则两个孩子(依大小排列)的性别为(男男)，(男女)，(女男)，(女女)的可能性是一样的。
- 记A为随机选取两娃家庭有一男一女的事件，则$P(A)=\frac{1}{2}$。
- 如果我们预先知道某个家庭至少有一个女孩(设为事件B)，那么，上述事件A的概率便应是$\frac{2}{3}$。

条件概率

【条件概率】设A和B是两事件且$P(B) > 0$，定义

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

为在已知事件B发生条件下事件A发生的概率。

乘法公式

大多数应用中，我们直接获取的是条件概率。

【乘法公式】$P(AB) = P(B)P(A\,|B)$

一般地，我们有

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

乘法公式练习

设某光学仪器厂制造的透镜，第一次落下时打破的概率为$\frac{1}{2}$，若第一次落下未打破，第二次落下打破的概率为$\frac{7}{10}$，若前两次落下未打破，第三次落下打破的概率为$\frac{9}{10}$。

试求透镜落下三次后未打破的概率。

乘法公式练习

解：设A_i为第i次落下打破透镜事件，B为三次落下未打破透镜，

则有

$$B=\overline{A_1}\cap\overline{A_2}\cap\overline{A_3}$$

则有

$$P(B)=P\left(\overline{A_1}\cap\overline{A_2}\cap\overline{A_3}\right)=P\left(\overline{A_3}\middle|\overline{A_1}\cap\overline{A_2}\right)P\left(\overline{A_2}\middle|\overline{A_1}\right)P\left(\overline{A_1}\right)$$

$$=\left(1-\frac{9}{10}\right)\left(1-\frac{7}{10}\right)\left(1-\frac{1}{2}\right)=0.015$$

完备事件组

【完备事件组】设 B_1， $\cdots$， B_n 是样本空间Ω的一个事件组，若满足

- $B_i \cap B_j = \emptyset$，对于任意$i \neq j$；
- $\cup_i B_i = \Omega$

则B_1， $\cdots$， B_n称为Ω的完备事件组。

全概率公式

【全概率公式】设B_1，⋯，B_n是样本空间S的一个完备事件组，且$P(B_i) > 0(i = 1，2，\cdots n)$，则对于任一随机事件$A$，有

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

全概率公式练习

一等小麦种子中混合2%的二等种子，1.5%三等种子，1%的四等种子。用一等、二等、三等、四等种子长出的穗含50颗以上麦粒的概率分别是0.5，0.15，0.1，0.05。

求这批种子所结的穗含有50颗以上麦粒的概率。

全概率公式练习

设从这批种子中任选一颗是一等、二等、三等、四等种子的事件分别记为A_1，A_2，A_3，A_4，则它们构成样本空间的一个完备事件组。设B表示在这批种子中所结的穗含有50颗以上麦粒的事件，则由全概率公式得

$$P(B)=\sum_i P(A_i)P(B|A_i)$$

$$=0.955*0.5+0.02*0.15+0.15*0.1+0.01*0.05=0.4825$$

全概率公式练习

考卷中一道选择题有4个答案，仅有一个是正确的，设一个学生知道正确答案或不知道而乱猜是等可能的。

如果这个学生答对了，求他确实知道正确答案的概率。

全概率公式练习

样本空间可以划分为事件A：知道正确答案，$\bar{A}$不知道正确答案。 以B表示学生答对事件，则$A \subset B$，$P(AB) = P(A) = \frac{1}{2}$，$P(B \mid A) = 1$而$P(B|\bar{A}) = \frac{1}{4}$。我们需要求$P(A|B)$。首先由全概率公式

$$P(B) = P(A)P(B|A) + P(\bar{A})P(B|\bar{A})$$

$$\frac{1}{2} * 1 + \frac{1}{2} * \frac{1}{4} = \frac{5}{8}$$

故

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{4}{5}$$

更多详情，请咨询课程助教。

全概率公式练习

- 一个盒子中有3个红球，2个白球，每次从袋中任取一只，观察其颜色后放回，并再放入一只与所取之球颜色相同的球，若从盒子中连续取球4次，试求第1、2次取得白球、第3、4次取得红球的概率。
- 有甲乙两个袋子，甲袋中有两个白球，1个红球，乙袋中有两个红球，一个白球．这六个球手感上不可区别．今从甲袋中任取一球放入乙袋，搅匀后再从乙袋中任取一球，问此球是红球的概率?

贝叶斯公式

【贝叶斯公式】设A，B为两事件, $P(B)>0$, 则

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

这不仅仅是一个公式……

贝叶斯公式计算

某电子设备制造厂所用的元件是由三家元件制造厂提供的。根据以往的记录有以下的数据：

制造厂	次品率	份额
1	0.02	0.15
2	0.01	0.80
3	0.03	0.05

设这三家工厂的产品在仓库中是均匀混合的，且无区别的标志。

(1)在仓库中随机地取一只元件，求它是次品的概率；

(2)在仓库中随机地取一只元件，若已知取到的是次品，为分析此次品出自何厂，需求出此次品由三家工厂生产的概率分别是多少。

贝叶斯公式计算

设A表示“取到的是一只次品”，$B_i(i=1，2，3)$ 表示“所取到的产品是由第i家工厂提供的”。 有$P(B_1)=0.15$，$P(B_2)=0.80$，$P(B_3)=0.05$；$P(A|B_1)=0.02$，$P(A|B_2)=0.01$，$P(A|B_3)=0.03$。

(1) 由全概率公式

$$P(A)=P(A|B_1)P(B_1)+P(A|B_2)P(B_2)+P(A|B_3)P(B_3)=0.0125$$

(2) 由贝叶斯公式

$$P(B_1|A)=\frac{P(A|B_1)P(B_1)}{P(A)}=0.24，P(B_2|A)=0.64，P(B_3|A)=0.12$$

贝叶斯公式用途

- 贝叶斯定理往往与全概率公式同时使用。全概率公式一用于“由因求果”问题，而贝叶斯定理一般用于“执果寻因”问题。
- 例如，在医疗诊断中，为了诊断到底是哪种疾病，对病人进行观察与检查，确定了某项身体指标 (如是体温、验血)，然后采用这类指标来帮助诊断。这时就可以用贝叶斯公式来计算有关概率。

贝叶斯公式计算

根据以往的临床记录，某种诊断癌症的试验具有如下的效果：若以A表示事件“反应为阳性”，以C表示事件“被诊断者患有癌症”，则有$P(A|C) = 0.95$，$P(\bar{A}|\bar{C})$=0.95。现在对某区域人群进行综合调查，预测患癌概率为0.005，即$P(C) = 0.005$。

假设某一天，你去医院检查反应为阳性，试问你患癌的概率为多少？即求$P(C|A)$。

贝叶斯公式计算

解：已知$P(A|C)=0.95$，$P(A|\bar{C})=1-P(\bar{A}|\bar{C})=0.05$，$P(C)=0.005$，$P(\bar{C})=0.995$，由贝叶斯公式

$$P(C|A)=\frac{P(A|C)P(C)}{P(A|C)P(C)+P(A|\bar{C})P(\bar{C})}=0.087$$

本题的结果表明，虽然$P(A|C)=0.95$，$P(\bar{A}|\bar{C})$=0.95，这两个概率都比较高。

然而$P(C|A)$ 只有0.087，即其试验正确性只有8.7%。

如果混淆$P(A|C)$和$P(C|A)$，将会得出错误的诊断。

贝叶斯计算练习

数字通讯过程中，信源发射0、1两种状态信号，其中发0的概率为0.55，发1的概率为0.45。由于信道中存在干扰，在发0的时候，接收端分别以概率0.9、0.05和0.05接收为0、1和“不清”。在发1的时候，接收端分别以概率0.85、0.05和0.1接收为1、0和“不清”。

现接收端接收到一个“1”的信号。问发射端发的是0的概率是多少?

独立性

设A，B是两个事件，如果满足等式

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

则称事件A和B相互独立，简称A和B独立。

独立性

【定理】设A，B是两个事件，且$P(A)>0$，若A，B相互独立，则

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

反之亦然。

【定理】若事件A和B相互独立，则下列各对事件：A与$\bar{B}$，B与$\bar{A}$，$\bar{A}$与$\bar{B}$，也相互独立。

独立性

设A，B，C是三个事件，如果满足等式

$$P(AB)=P(A)P(B),\quad P(BC)=P(B)P(C),\quad P(AC)=P(A)P(C)$$

$$P(ABC)=P(A)P(B)P(C)$$

则称事件A，B，C相互独立。

更多详情，请咨询课程助教。