1.5　全概率公式与贝叶斯公式

一、全概率公式

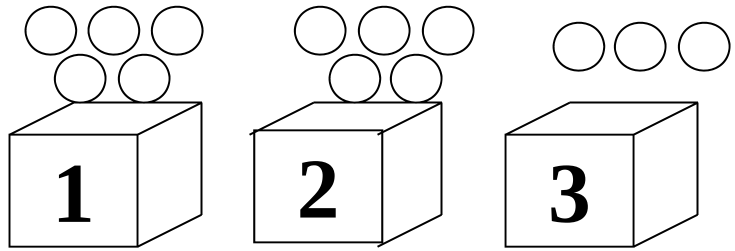
二、贝叶斯公式

三、小结

一、全概率公式

先看一个例子:

有三个箱子,分别编号为1,2,3.1号箱装有1个红球4个白球,2号箱装有2红3白球 , 3号箱装有3 红球. 某人从三箱中任取一箱,从中任意摸出一球,求取得红球的概率.



解　记 A_i={球取自i号箱}, i=1,2,3;

B ={取得红球}

其中　A_1、A_2、A_3两两互斥

B发生总是伴随着A_1，A_2，A_3 之一同时发生,

即 $B=A_1B+A_2B+A_3B$，

且 A_1B、A_2B、A_3B 两两互斥

运用加法公式得到

$$P(B)=P(A_1B)+P(A_2B)+P(A_3B)$$

对求和中的每一项运用乘法公式得

$$P(B)=\sum_{i=1}^{3}P(A_i)P(B|A_i)$$

代入数据计算得：$P(B)=8/15$

将此例中所用的方法推广到一般的情形，就得到在概率计算中常用的全概率公式.

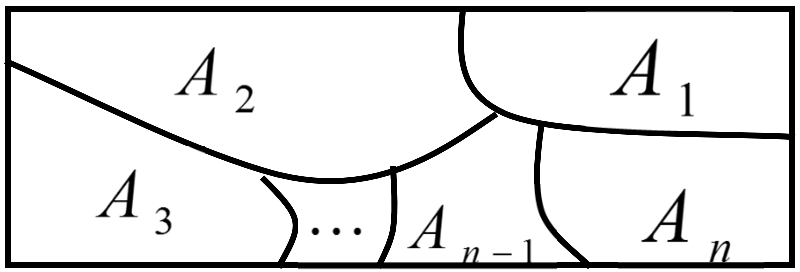
1. 样本空间的划分

定义 设 Ω 为试验E的样本空间, $A_1, A_2, \cdots, A_n$ 为 E 的一组事件, 若

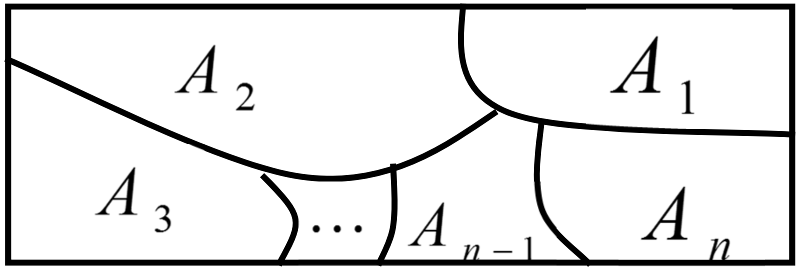
$1^0 \quad A_i A_j = \varnothing, i, j = 1,2,\cdots,n;$

$2^0 \quad A_1 \cup A_2 \cup \cdots \cup A_n = \Omega,$

则称 $A_1, A_2, \cdots, A_n$ 为样本空间 Ω 的一个划分.



或称$A_1,A_2,\ldots,A_n$构成一个完备事件组.



注意：若$A_1,A_2,\ldots,A_n$是样本空间的一个划分, 那么,对于每次试验, 事件$A_1,A_2,\ldots,A_n$中必有一个且仅有一个发生.

2. 全概率公式

定义 设Ω为试验E的样本空间,B为E的事件,$A_1, A_2, \cdots, A_n$为Ω的一个划分,且$P(A_i) > 0$ $(i = 1,2,\cdots,n)$,则

$$\boxed{\begin{aligned} P(B) &= P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) \\ &\quad + \cdots + P(B \mid A_n)P(A_n) \\ &= \sum_{i=1}^{n} P(A_i)P(B \mid A_i) \end{aligned}}$$

全概率公式

证明 $B = B\Omega = B\bigcap(A_1 \bigcup A_2 \bigcup \cdots A_n)$

$$= BA_1 \bigcup BA_2 \bigcup \cdots \bigcup BA_n.$$

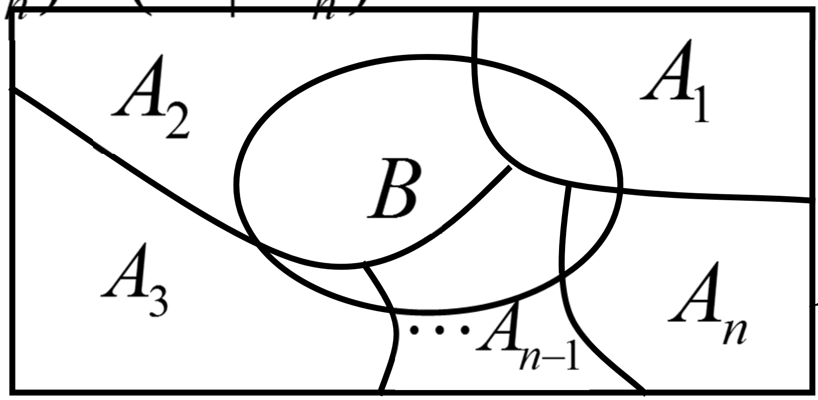
由 $A_iA_j = \varnothing \Rightarrow (BA_i)(BA_j) = \varnothing$

$$\Rightarrow P(B) = P(BA_1) + P(BA_2) + \cdots + P(BA_n)$$

$$\Rightarrow P(B) = P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + \cdots$$

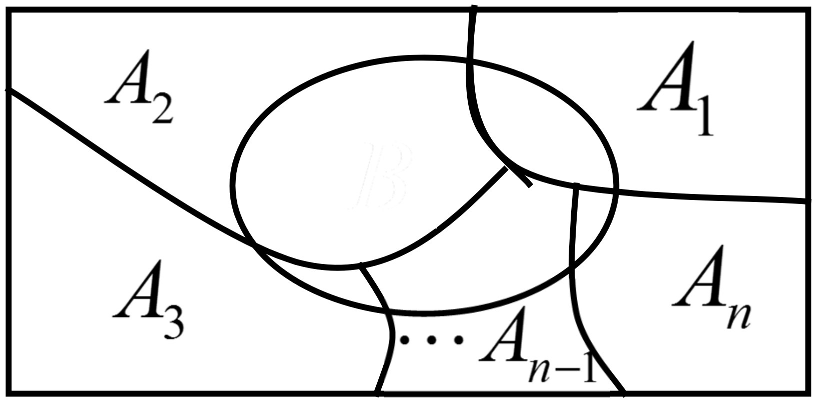
$$+ P(A_n)P(B \mid A_n)$$

图示



化整为零
各个击破

说明　全概率公式的主要用途在于它可以将一个复杂事件的概率计算问题,分解为若干个简单事件的概率计算问题,而这些简单事件组成一个互不相容事件组，使得某个未知事件与这组互不相容事件中至少一个同时发生，最后应用概率的可加性求出最终结果.



我们还可以从另一个角度去理解 全概率公式.

某一事件A的发生有各种可能的原因 ，如果A是由原因B_i (i=1,2,…,n) 所引起，则A发生的概率是

$$P(AB_i)=P(B_i)P(A\,|B_i)$$

每一原因都可能导致A发生，故A发生的概率是各原因引起A发生概率的总和，即全概率公式.

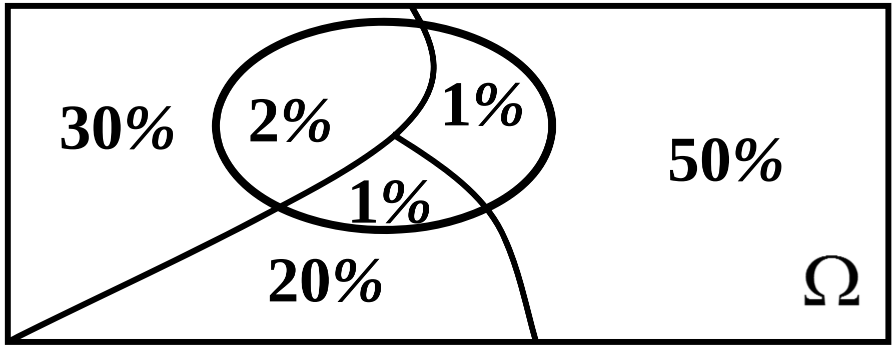
由此可以形象地把全概率公式看成为“由原因推结果”，每个原因对结果的发生有一定的“作用”，即结果发生的可能性与各种原因的“作用”大小有关. 全概率公式表达了它们之间的关系 .

例1 有一批同一型号的产品,已知其中由一厂生产的占 **30%** , 二厂生产的占 **50%** , 三厂生产的占 **20%**, 又知这三个厂的产品次品率分别为**2%** , **1%**, **1%**,问从这批产品中任取一件是次品的概率是多少?

解 设事件 A 为“任取一件为次品”,

事件 B_i 为"任取一件为 i 厂的产品", $i = 1,2,3.$

$$B_1 \cup B_2 \cup B_3 = \Omega, \quad B_i B_j = \varnothing, \quad i, j = 1,2,3.$$



由全概率公式得

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

$$P(B_1)=0.3,\ \ P(B_2)=0.5,\ \ P(B_3)=0.2,$$

$$P(A|B_1)=0.02,\ P(A|B_2)=0.01,\ P(A|B_3)=0.01,$$

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

$$=0.02\times0.3+0.01\times0.5+0.01\times0.2=0.013.$$

例2 墨西哥湾和美国的东部海域每年都会发生飓风，大多是发生在夏季和秋季。风速不小于75mph（每小时75英里）时被定义为飓风，飓风可分为从C1到C5五个类别。显然，随着级别增加，飓风的发生频度是减少的。例如，持续风速达到大于等于150mph的C5类飓风事实上很少发生。假设每年至多只会有一次飓风袭击沿墨西哥湾的美国路易斯安那州南部海岸的某给定地区。

不同类别的飓风的年发生概率为

$P(C1)=0.35; P(C2)=0.25; P(C3)=0.14; P(C4)=0.05; P(C5)=0.01.$

不同类别的飓风的年发生概率为

$P(C1)=0.35; P(C2)=0.25; P(C3)=0.14; P(C4)=0.05; P(C5)=0.01.$

在该地区域，某建筑是否发生结构破坏，取决于该建筑可能遭受的飓风类别。假设该建筑破坏发生的条件概率为：

$P(D|C1)=0.05; P(D|C2)=0.10; P(D|C3)=0.25;$

$P(D|C4)=0.60; P(D|C5)=1.00; P(D|C0)=0.$

那么该建筑遭受风灾破坏的年概率为多少？年破坏率的主要贡献来自于哪类飓风？

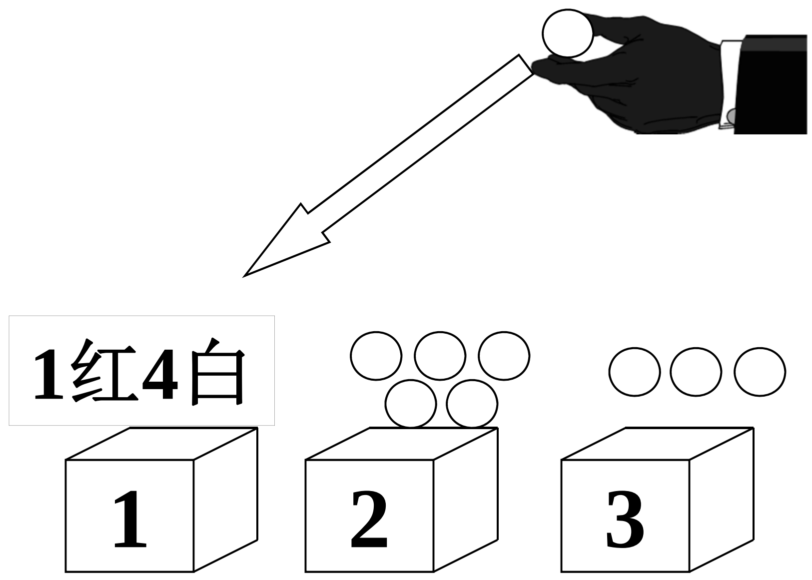
3. 贝叶斯公式

再看一个例子:

某人从任一箱中任意摸出一球，发现是红球,求该球是取自1号箱的概率.

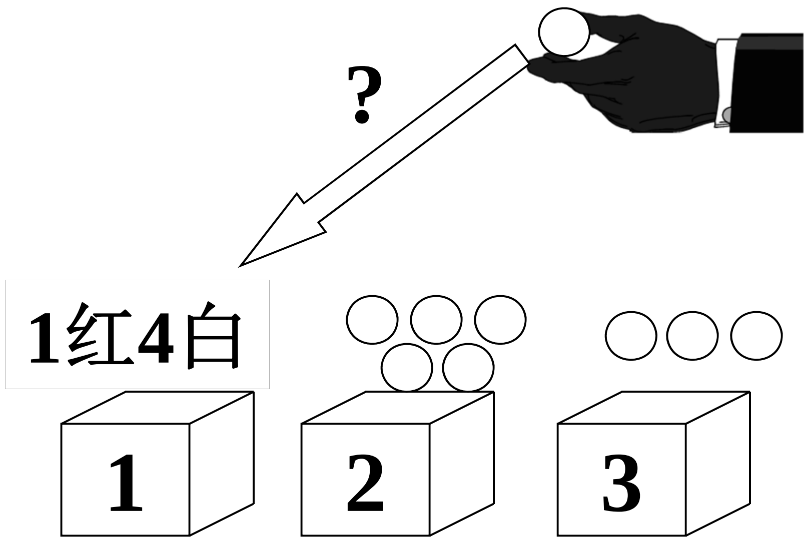
或者问:

该球取自哪号箱的可能性最大?

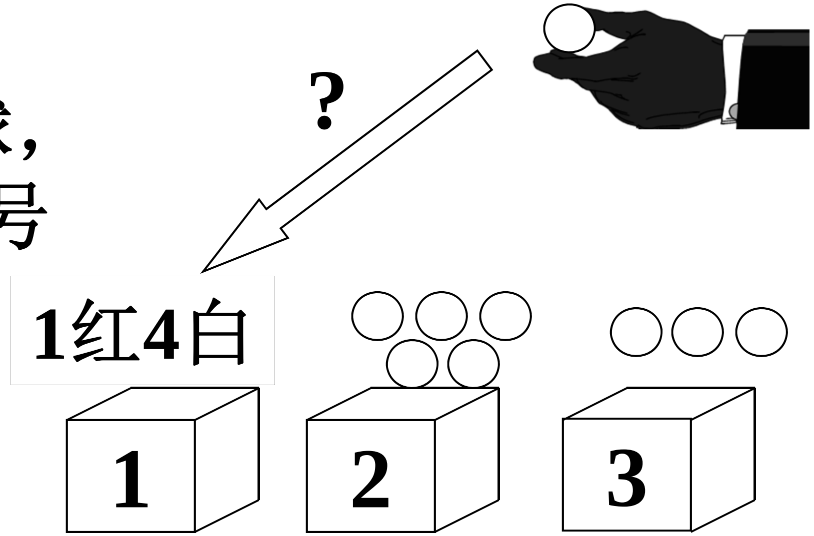


这一类问题是“已知结果求原因”. 在实际中更为常见，它所求的是条件概率，是已知某结果发生条件下，探求各原因发生可能性大小.

有三个箱子，分别编号为1,2,3，1号箱装有1个红球4个白球，2号箱装有2红球3白球，3号箱装有3红球. 某人从三箱中任取一箱，从中任意摸出一球，发现是红球,求该球是取自1号箱的概率 .



某人从任一箱中任意摸出一球，发现是红球，求该球是取自1号箱的概率.



记 B_i={球取自i号箱}, i=1,2,3;

A ={取得红球}

求$P(B_1|A)$

$$P(B_1 \mid A) = \frac{P(B_1 A)}{P(A)} = \frac{P(B_1)P(A \mid B_1)}{\sum_{k=1}^{3} P(B_k)P(A \mid B_k)}$$

运用全概率公式计算$P(A)$

将这里得到的公式一般化，就得到 贝叶斯公式

定义 设Ω为试验E的样本空间，B为E的事件，$A_1, A_2, \cdots, A_n$为Ω的一个划分，且$P(B) > 0$，$P(A_i) > 0 (i = 1,2,\cdots,n)$，则

$$\boxed{P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{\sum_{j=1}^{n} P(B \mid A_j)P(A_j)}, \quad i = 1,2,\cdots,n.}$$

称此为**贝叶斯公式**.

证明

$$\mathbf{P(A_i|B) = \frac{P(B\,|\,A_i)P(A_i)}{P(B)}}$$

$$\mathbf{= \frac{P(A_i)P(B\,|\,A_i)}{\sum\limits_{j=1}^{n} P(A_j)P(B\,|\,A_j)}} \qquad i = 1,2,\cdots,n.$$

[证毕]

特别在n=2时, 并将A_1记为B,此时

B_2就是$\overline{B}$,那么,全概率公式和贝叶斯公式分别成为

$$P(B)=P(B\mid A)P(A)+P(B\mid \overline{A})P(\overline{A}),$$

$$P(A\mid B)=\frac{P(B\mid A)P(A)}{P(B\mid A)P(A)+P(B\mid \overline{A})P(\overline{A})}$$

例3 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:

元件制造厂	次品率	提供元件的份额
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,求此次品出由三家工厂生产的概率分别是多少.

解 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".

则 B_1, B_2, B_3 是样本空间 Ω 的一个划分,

且 $P(B_1)=0.15,\quad P(B_2)=0.80,\quad P(B_3)=0.05,$

$$P(A|B_1)=0.02,\quad P(A|B_2)=0.01,\quad P(A|B_3)=0.03.$$

(1) 由全概率公式得

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

$$=0.0125.$$

(2) 由贝叶斯公式得

$$P(B_1|A)=\frac{P(A|B_1)P(B_1)}{P(A)}=\frac{0.02\times 0.15}{0.0125}=0.24.$$

$$P(B_2|A)=\frac{P(A|B_2)P(B_2)}{P(A)}=0.64,$$

$$P(B_3|A)=\frac{P(A|B_3)P(B_3)}{P(A)}=0.12.$$

故这只次品来自第 2 家工厂的可能性最大 .

例4　　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"试验反应为阳性",以 C 表示事件"被诊断者患有癌症",则有 $P(A|C)=0.95$, $P(\overline{A}|\overline{C})=0.95$.现在对自然人群进行普查,设被试验的人患有癌症的概率为0.005,即 $P(C)=0.005$,试求 $P(C|A)$.

解　因为　$P(A|C)=0.95$,

$$P(A|\overline{C})=1-P(\overline{A}|\overline{C})=0.05,$$

$$P(C)=0.005,\quad P(\overline{C})=0.995,$$

由贝叶斯公式得所求概率为

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

$$=0.087.$$

即平均**1000**个具有阳性反应的人中大约只有**87**人患有癌症.

例5 对以往数据分析结果表 明,当机器调整得良好时,产品的合格率为 98%,而当机器发生某种故障时,其合格率为 55%.每天早上机器开动时,机器调整良好的概率为 95%.试求已知某日早上第一件产品是合格 时,机器调整得良好的概率是多少?

解 设 A 为事件"产品合格".

B 为事件"机器调整良好".

则有

$$P(A|B) = 0.98, \quad P(A|\overline{B}) = 0.55,$$

$$P(B)=0.95,\qquad P(\overline{B})=0.05,$$

由贝叶斯公式得所求概率为

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

$$=\frac{0.98\times0.95}{0.98\times0.95+0.55\times0.05}=0.97.$$

即当生产出第一件产品是合格品时,此时机器调整良好的概率为0.97.

先验概率与后验概率

上题中概率 **0.95** 是由以往的数据分析得到的, 叫做先验概率.

而在得到信息之后再重新加以修正的概率 **0.97** 叫做后验概率.

三、小结

1.条件概率 $P(B|A)=\dfrac{P(AB)}{P(A)}$ $\longrightarrow$ 乘法定理

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

$\downarrow$

全概率公式

$$P(A)=P(B_1)P(A|B_1)+P(B_2)P(A|B_2)+\cdots+P(B_n)P(A|B_n)$$

$\downarrow$

贝叶斯公式

$$P(B_i|A)=\frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{n}P(B_j)P(A|B_j)},\ i=1,2,\cdots,n.$$

备份题

例1 设袋中有**4**只白球**, 2**只红球 **,**

(1) 无放回随机地抽取两次**,** 每次取一球**,** 求在两次抽取中至多抽到一个红球的概率**?**

(2) 若无放回的抽取 **3**次**,** 每次抽取一球**,** 求

(*a*) 第一次是白球的情况下**,** 第二次与第三次均是白球的概率**?**

(*b*) 第一次与第二次均是白球的情况下，第三次是白球的概率**?**

解　(1)设 A 为事件 "两次抽取中至多抽到一 个红球" 事件 A_1 为"第一次抽取到红球 "A_2 为"第二次抽取到红球 ".

则有　$A=\overline{A_1}\,\overline{A_2}+\overline{A_1}A_2+A_1\overline{A_2}$,

$$P(A)=P(\overline{A_1}\,\overline{A_2})+P(\overline{A_1}A_2)+P(A_1\overline{A_2})$$

$$=P(\overline{A_1})P(\overline{A_2}|\overline{A_1})+P(\overline{A_1})P(A_2|\overline{A_1})+P(A_1)P(\overline{A_2}|A_1)$$

$$=\frac{4}{6}\cdot\frac{3}{5}+\frac{4}{6}\cdot\frac{2}{5}+\frac{2}{6}\cdot\frac{4}{5}=\frac{14}{15}.$$

(2)设事件 A_i 为"第 i 次取出的是白球", $i = 1,2,3$.

$$(a)\ P(A_2A_3|A_1) = \frac{P(A_1A_2A_3)}{P(A_1)},$$

因为 $P(A_1) = \frac{4}{6} = \frac{2}{3}$, $P(A_1A_2A_3) = \binom{4}{3} \Big/ \binom{6}{3} = \frac{1}{5}$,

所以 $P(A_2A_3|A_1) = \frac{P(A_1A_2A_3)}{P(A_1)} = \frac{1/5}{2/3} = \frac{3}{10}$.

(*b*) $P(A_3|A_1A_2)=\dfrac{P(A_1A_2A_3)}{P(A_1A_2)}$,

因为 $P(A_1A_2)=\dbinom{4}{2}\Big/\dbinom{6}{2}=\dfrac{2}{5}$,

$$P(A_1A_2A_3)=\binom{4}{3}\Big/\binom{6}{3}=\frac{1}{5},$$

所以 $P(A_3|A_1A_2)=\dfrac{P(A_1A_2A_3)}{P(A_1A_2)}=\dfrac{1/5}{2/5}=\dfrac{1}{2}$.

掷骰子试验

例2 掷两颗骰子, 已知两颗骰子点数之和为7, 求其中有一颗为1点的概率.

解 设事件A为“两颗点数之和为 7”, 事件 B 为“一颗点数为1”.

两颗点数之和为 7 的种数为 3,

其中有一颗为 1 点的种数为 1,

故所求概率为 $P = \frac{1}{3}.$

例3 设一仓库中有10 箱同种规格的产品, 其中由甲、乙、丙三厂生产的分别有5箱 , 3箱, 2 箱, 三厂产品的废品率依次为 0.1, 0.2, 0.3 从这 10 箱产品中任取一箱 , 再从这箱中任取一件产品, 求取得的正品概率.

解 设 A 为事件"取得的产品为正品"B_1, B_2, B_3,

分别表示"任取一件产品是甲、乙、丙生产的",

由题设知 $P(B_1) = \frac{5}{10}$, $P(B_2) = \frac{3}{10}$, $P(B_3) = \frac{2}{10}$.

$$P(A|B_1) = 0.9,\ \ P(A|B_2) = 0.8,\ \ P(A|B_3) = 0.7,$$

故

$$P(A) = \sum_{i=1}^{3} P(B_i)P(A|B_i)$$

$$= \frac{5}{10}\cdot\frac{9}{10} + \frac{3}{10}\cdot\frac{8}{10} + \frac{2}{10}\cdot\frac{7}{10}$$

$$= 0.82.$$