# 1.1 排列与组合

## 1.1.1 排列

从$n$个不同元素中任取$r(r \le n)$个元素排成一排（不能重复选择元素，要考虑元素的先后顺序），称为一个排列（Permutation）。按乘法原理，此种排列共有

$$n \times (n-1) \times \cdots \times (n-r+1)$$

种，记作$P_n^r$，可以读作：

$$n \text{ Pick } r$$

若$r = n$，称为全排列，全排列数共有$n!$个，记为$P_n$。

## 1.1.2 组合

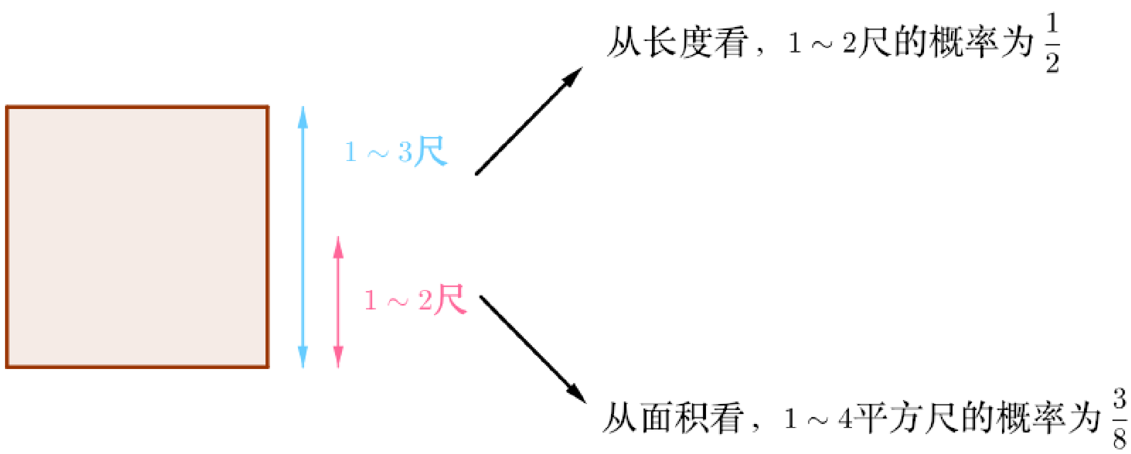
从$n$个不同元素中任取$r(r \le n)$个元素并成一组(不区分顺序)，称为一个组合（Combination），组合总数为：

$$C_n^r = \binom{n}{r} = \frac{P_n^r}{r!} = \frac{n!}{r!(n-r)!}$$

可以读作：

$$n \text{ Choose } r$$

## 1.3 几何概型




通过这个典型的例子可以看到，样本空间选择不同，同一个问题会得到了不同的概率。值得注意的是，就数学而言，这两个结论都是对的；在现实中哪个是对的，还需要通过采样、测量来决定。

# 1.2 事件的关系与运算

### 1.2.1 关系

| | |
|---|---|
| 包含 | $A \subseteq B \quad \text{或} \quad B \subseteq A$ |
| 相等 | $A \subseteq B \quad \text{且} \quad B \subseteq A$ |
| 互斥 | $A \cap B = \varnothing$ |
| 对立 | $A = \overline{B}$ |

#### 1.2.1.1 包含

设$A$、$B$为两个事件，若$A \subseteq B$，则可推出：

$$\begin{aligned} P(B-A) = P(B) - P(A) \qquad (1) \\ P(A) \leq P(B) \qquad (2) \end{aligned}$$

#### 1.2.1.2 加法公式

对于任意两个事件，有：

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

也常写作如下的形式：

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

所以这个公式也称为概率的加法公式。

两个事件加法公式也可以推广至三个事件：

$$\begin{aligned} P(A \cup B \cup C) = & P(A) + P(B) + P(C) \\ & \underbrace{-P(A \cap B) - P(A \cap C) - P(B \cap C)}_{\text{重复计算的部分需要“排斥”出去}} \\ & \underbrace{+P(A \cap B \cap C)}_{\text{多“排斥”出去的部分需要“包容”回来}} \end{aligned}$$

所以，加法公式也被称作容斥定理（Inclusion–exclusion principle）。

### 1.2.2 运算

| | 定义 | 类比 |
| --- | --- | --- |
| 并 | $A \cup B = \{x \mid x \in A \text{ 或 } x \in B\}$ | $+$ |
| 交 | $A \cap B = \{x \mid x \in A \text{ 且 } x \in B\}$ | $\times$ |
| 差 | $A - B = \{x \mid x \in A \text{ 且 } x \notin B\}$ | $-$ |
| 补 | $\overline{A} = B \iff B = \Omega - A$ | |

然后这些基本运算和实数基本运算的性质差不多。

#### 1.2.2.1 交换律

- $A + B = B + A$
- $AB = BA$

#### 1.2.2.2 结合律

- $A + B + C = A + (B + C)$
- $ABC = A(BC)$

#### 1.2.2.3 分配律

- $(A + B)C = AC + BC$
- $(A - B)C = AC - BC$
- $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$

#### 1.2.2.4 德摩根定律

- $\overline{A \cup B} = \overline{A} \cap \overline{B}$
- $\overline{A \cap B} = \overline{A} \cup \overline{B}$

## 1.2 条件概率

### 1.2.1 定义

设$A$和$B$是样本空间$\Omega$中的两事件，若$P(B) > 0$，则称：

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

为“假设条件为$B$时的$A$的概率”，简称条件概率。也常写作：

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

概率与条件概率两者最重要的区别是：

- $P(A)$是以$\Omega$为样本空间，这是一个全局的空间，包含所有可能性
- $P(A|B)$是以$B$为样本空间，相对而言是一个局部空间，有些可能性被排除掉了

### 1.2.2 性质

条件概率虽然是被推广出来的定义，但这种推广也是很合理的，一方面符合过往的经验（可以用古典派进行验算），另外一方面符合概率三大公理：

对于条件概率$P(A|B)$，满足以下三大公理：

（1）**非负性公理**：

$$P(A|B)\geq 0$$

（2）**规范性公理**：

$$P(\Omega|B)=1$$

（3）**可加性公理**：设$A_1$、$A_2$、$\cdots$为两两不相容的事件，即$A_i\cap A_j=\varnothing$（$i\neq j$），有：

$$P\left(\bigcup_{n=1}^{\infty}A_n|B\right)=\sum_{n=1}^{\infty}P\left(A_n|B\right)$$

证明了条件概率是数学意义上的概率之后，马上就知道了，之前所做的推论在条件概率上也成立：

| | 概率 | 条件概率 |
|---|---|---|
| 空集 | $P(\varnothing)=0$ | $P(\varnothing\|C)=0$ |
| 包含<br>$(A\subseteq B)$ | $P(B-A)=P(B)-P(A)$<br>$P(A)\leq P(B)$ | $P(B-A\|C)=P(B\|C)-P(A\|C)$<br>$P(A\|C)\leq P(B\|C)$ |
| 加法 | $P(A+B)=P(A)+P(B)-P(AB)$ | $P(A+B\|C)=P(A\|C)+P(B\|C)-P(AB\|C)$ |

这里也可看出把概率公理化的好处，从很简单的三个公理开始，对各种结论进行证明。反过来，如果某定义满足这三个公理，那么概率论中的所有结论都对它成立。

### 1.2.3 乘法公式

从条件概率的定义我们马上可以推出乘法公式:

(1) 若$P(B)>0$，则:

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

(2) 若$P(A)>0$，则:

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

(3) 若$P(A_1\cdots A_n)>0$，则:

$$P(A_1\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1\cdots A_{n-1})$$

查看详细

(3) 式有点难记，其实还是有规律的。也可以手动来推，只需要反复运用条件概率的定义即可:

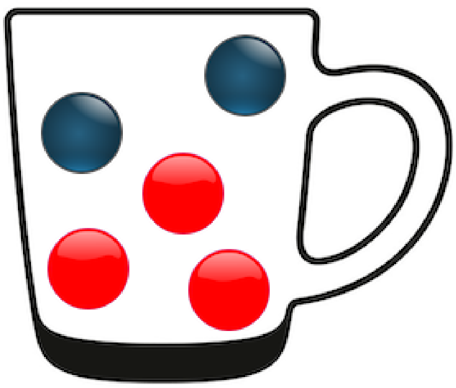
$$P(A_1A_2A_3)=P(A_1A_2)P(A_3|A_1A_2)$$

$$=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)$$

### 1.2.4 树形图

**树形图**有助于进一步理解条件概率。

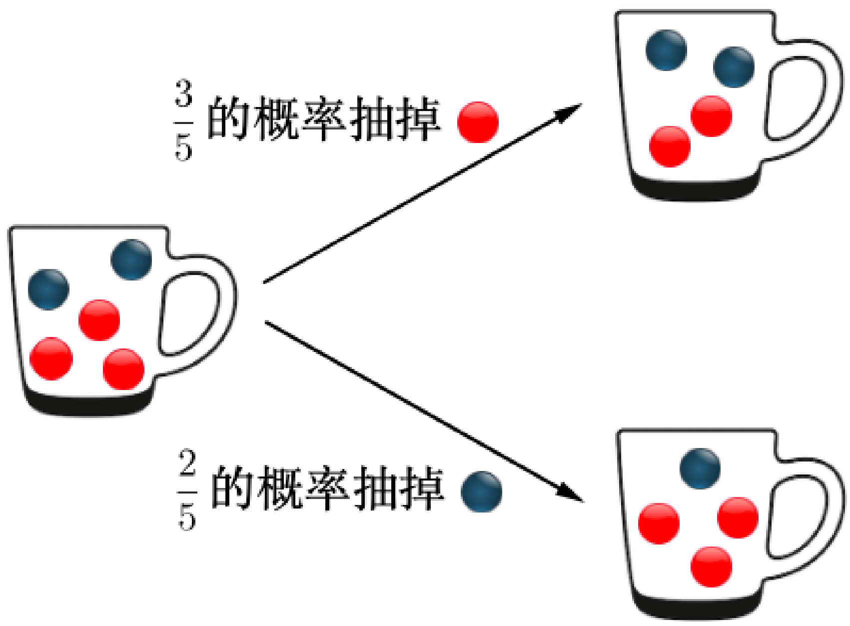
从一道题开始，杯中有三个红球、两个蓝球:



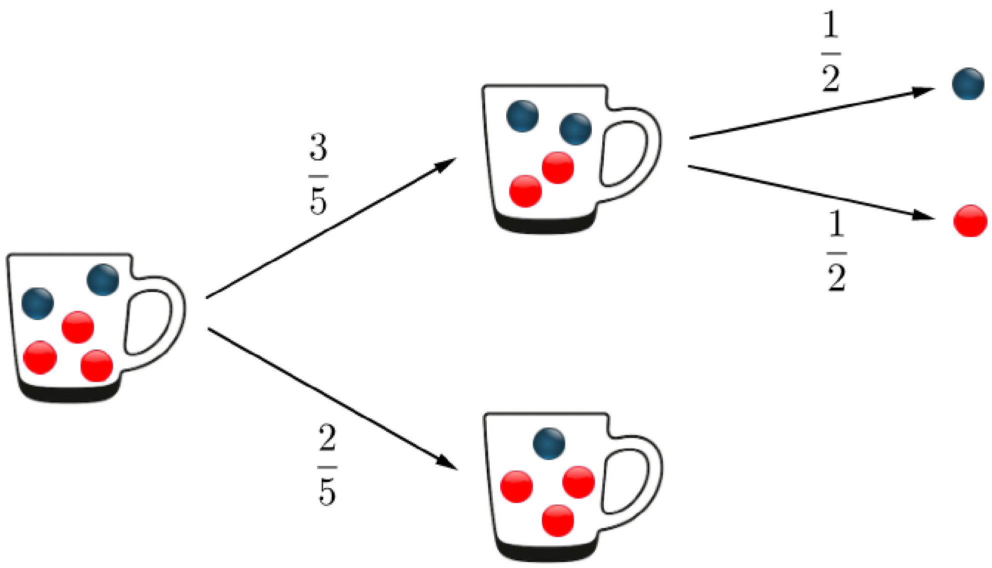
从中随机抽掉一个球后，再抽取一个蓝球的概率有多大?

**1 树形图**

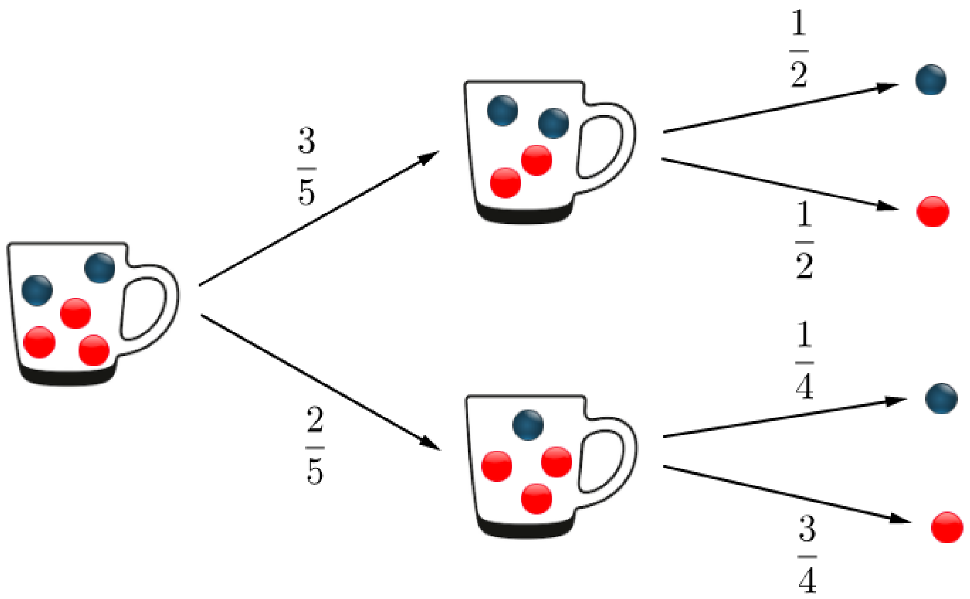
首先，从中随机抽掉一个球，有$\frac{3}{5}$的可能性抽掉红球、$\frac{2}{5}$的可能性抽掉蓝球:




如果抽掉红球，再抽取一个蓝球的概率为$\frac{1}{2}$：



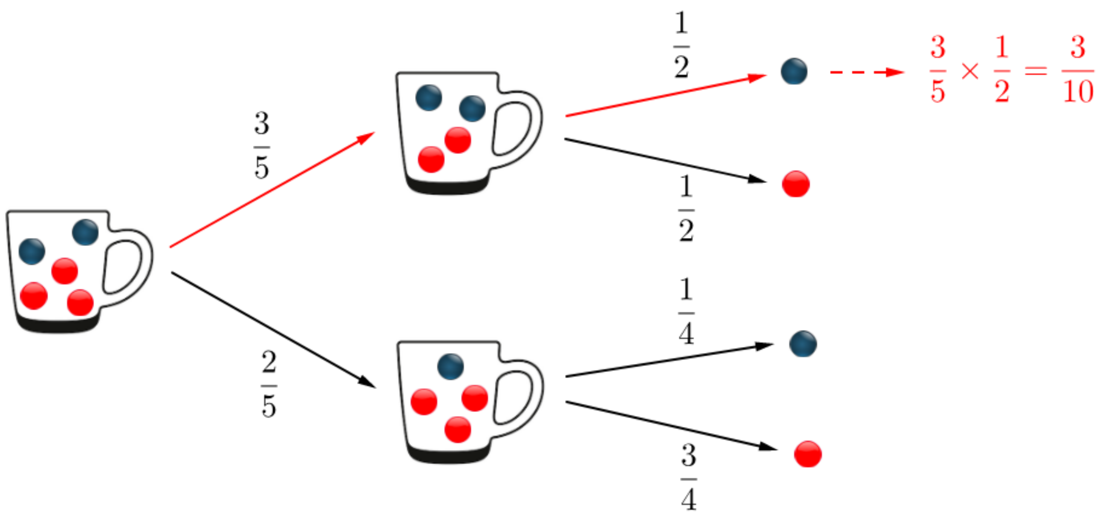

如果抽掉蓝球，再抽取一个蓝球的概率为$\frac{1}{4}$：



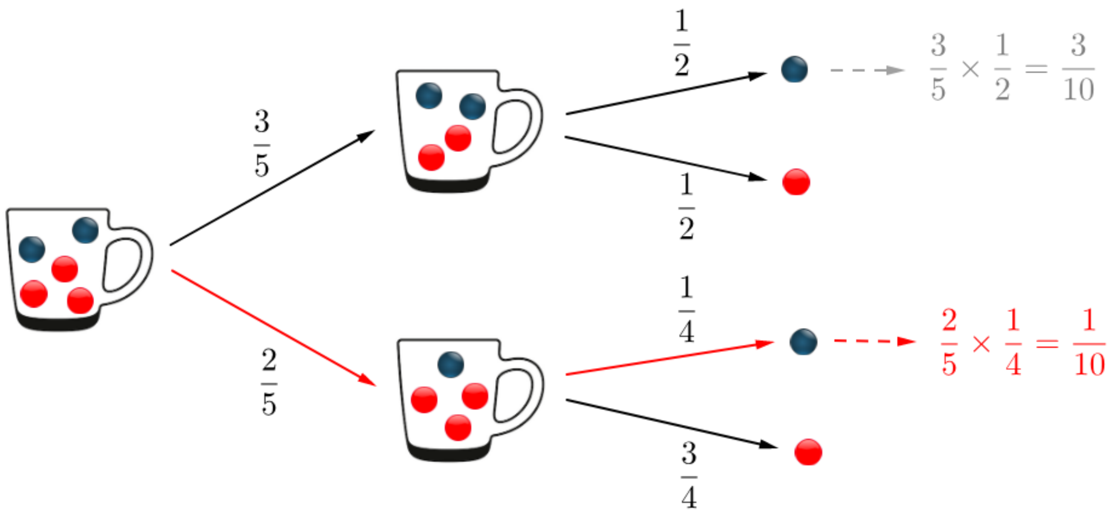

这就是树形图，下面可以用此来解题了。

**2 解题**

抽掉红球之后，再抽取一个蓝球的概率，可以沿着最上面的支线来计算（高中就是这么计算的，关于这点下一节会详细解释）：




而抽掉蓝球之后，再抽取一个蓝球的概率，可以沿着中间的支线来计算：

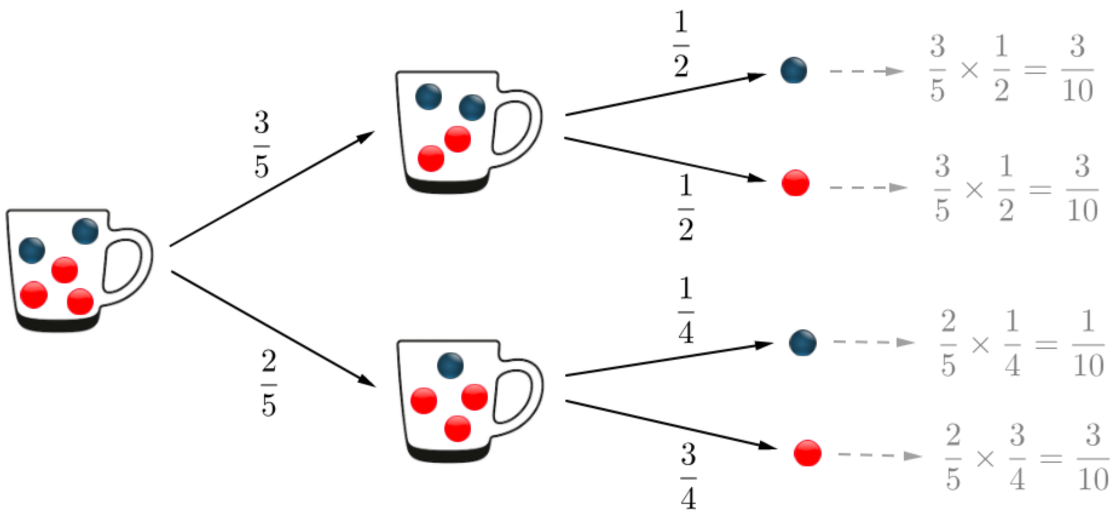



这样，随机抽掉一个球后，再抽取一个蓝球的概率是两条直线之和：

$$\frac{3}{10}+\frac{1}{10}=\frac{4}{10}$$

**3 验算**

如果树形图穷尽了所有的情况，且计算没有出错，那所有的分支概率之和应该为1；如果加起来不为1，说明有错漏，这个特点可以用来验算。来看看本题的所有分支：

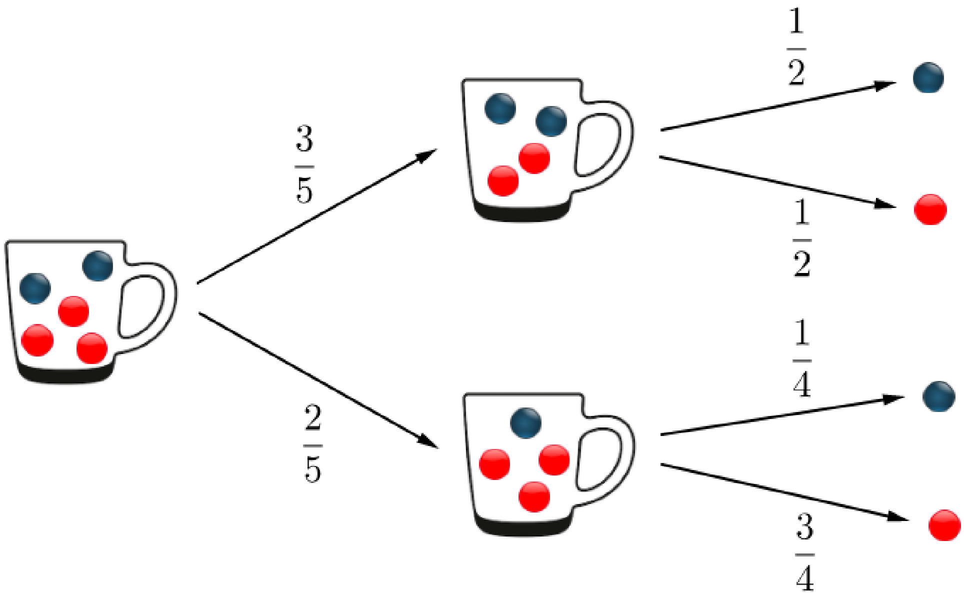



所有分支的概率之和确实为1：

$$\frac{3}{10}+\frac{3}{10}+\frac{1}{10}+\frac{3}{10}=1$$

**1.2.4.1 树形图与条件概率**

看着刚才那道题的树形图：

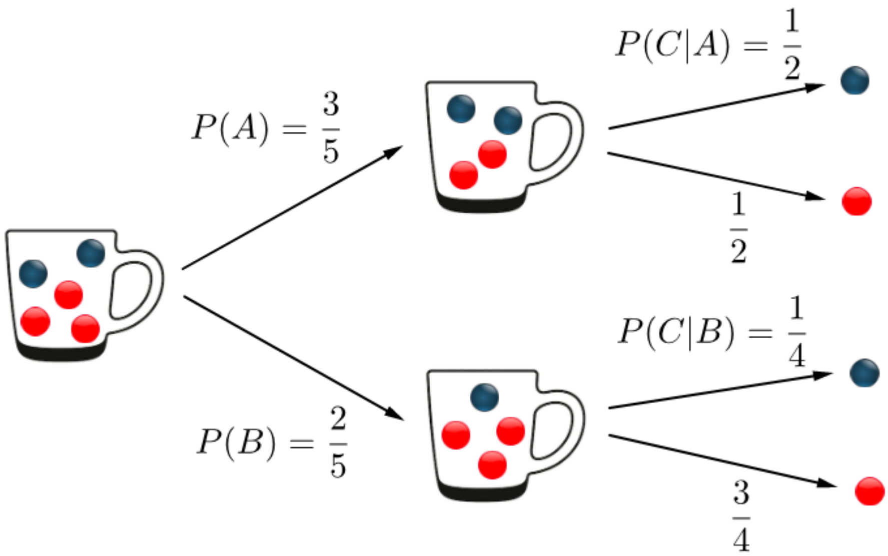



把这道题用代数来翻译下：

| | |
|---|---|
| 抽掉红球的概率为$\frac{3}{5}$ | $P(A)=\frac{3}{5}$ |
| 抽掉蓝球的概率为$\frac{2}{5}$ | $P(B)=\frac{2}{5}$ |
| 抽掉红球后抽取蓝球的概率为$\frac{1}{2}$ | $P(C\|A)=\frac{1}{2}$ |
| 抽掉蓝球后抽取蓝球的概率为$\frac{1}{4}$ | $P(C\|B)=\frac{1}{4}$ |
| 求，抽取蓝球的概率为多少？ | $P(C)=?$ |

查看详细

其实这是一个条件概率的题，把这些概率符号标注到图上：

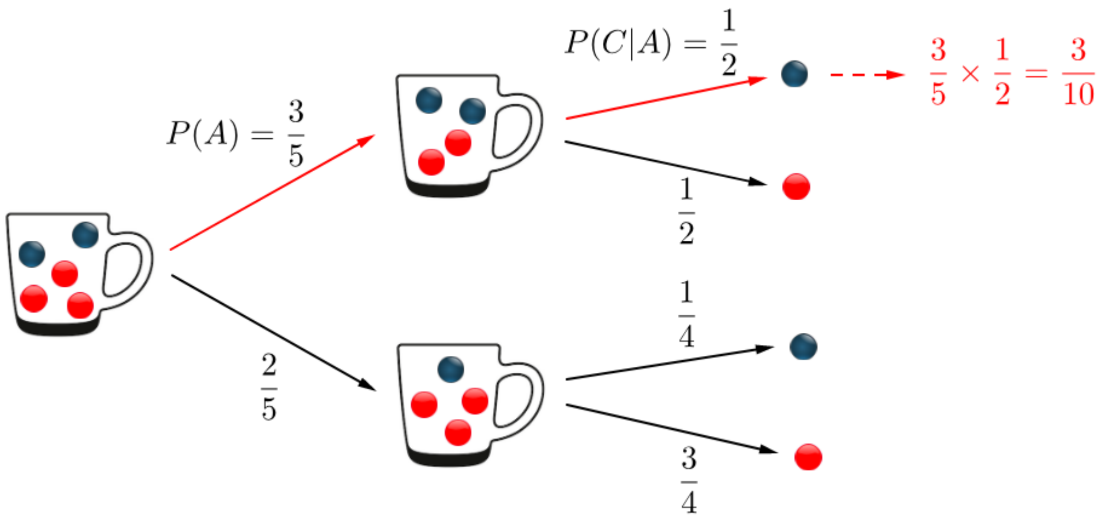



这里也可以看出条件概率的样本空间是局部样本空间：

| 样本空间 | 概率 或 条件概率 |
|---|---|
| $\Omega$ =  | $P(A)=\frac{3}{5}$ $\quad$ $P(B)=\frac{2}{5}$ |
| $A$ =  | $P(C\|A)=\frac{1}{2}$ |
| $B$ =  | $P(C\|B)=\frac{1}{4}$ |

#### 1.2.4.2 树形图与乘法公式

刚才求的抽掉红球后，再抽取一个蓝球的概率：




其实就是乘法公式的一次应用：

$$P(AC)=P(A)P(C|A)=\frac{3}{5}\times\frac{1}{2}=\frac{3}{10}$$

> 例题
>
> 1 假设生男、生女的概率均为 $\frac{1}{2}$,已知一对夫妇有两个孩子，现在知道他们有一个男孩，则另一个孩子为女孩的概率为多少
>
> 根据古典概率模型，两个孩子总共有四个等可能的情况：

$$\Omega = \{男女，男男，女男，女女\}$$

设事件$A$为有一个孩子为女孩，$B$为有一个孩子为男孩,则

$$A = \{男女, 女男, 女女\}, B = \{男女, 男男, 女男\}$$

$AB$即为有一个男孩一个女孩

$$A \cap B = \{男女, 女男\}$$

由此可得

$$P(B) = \frac{|B|}{|\Omega|} = \frac{3}{4} \quad P(AB) = \frac{|A \cap B|}{|\Omega|} = \frac{1}{2}$$

在已知一个是男孩的情况下，根据条件概率

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{2}{3}$$

**2** 已知一对夫妇有两个孩子，现在知道老大为男孩，则老二为女孩的概率为多少

样本空间$\Omega$为:

$$\Omega = \{(男, 女)，(男, 男), (女, 男), (女, 女)\}$$

设事件$A$="第一个孩子是男孩"，事件$B$="第二个孩子是女孩"

$$A = \{(男, 男), (男, 女)\}, B = \{(男, 女), (女, 女)\}$$

因此:

$$A \cap B = \{(男, 女)\}$$

根据条件概率:

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{\frac{|AB|}{|\Omega|}}{\frac{|A|}{|\Omega|}} = \frac{|AB|}{|A|} = \frac{1}{2}$$

**3** 袋中装有4个黑球，4个白球，从中随机取出一个球记下颜色后放回

再往袋中加入与取出的球同色的球2个，搅拌均匀后按同样的方式去求并加入求，一共重复三次

求三次中恰好有两次取出的球为黑球的概率$P_2$为

记第$i$次取到黑球为$A_i$,取到白球为$B_i$,其中$i = 1, 2, 3$,则

$$P(A_1) = P(A_2) = \frac{1}{2}$$

在第一次取出黑球的情况下，第二次取出黑球的概率即从6黑4白中取出黑球的概率，所以

$$P(A_2|A_1) = \frac{6}{10}$$

每个条件都可以用这样的方式计算，由互斥事件的概率加法公式知

$$P_2 = P(A_1 A_2 B_3) + P(A_1 B_2 A_3) + P(B_1 A_2 A_3)$$

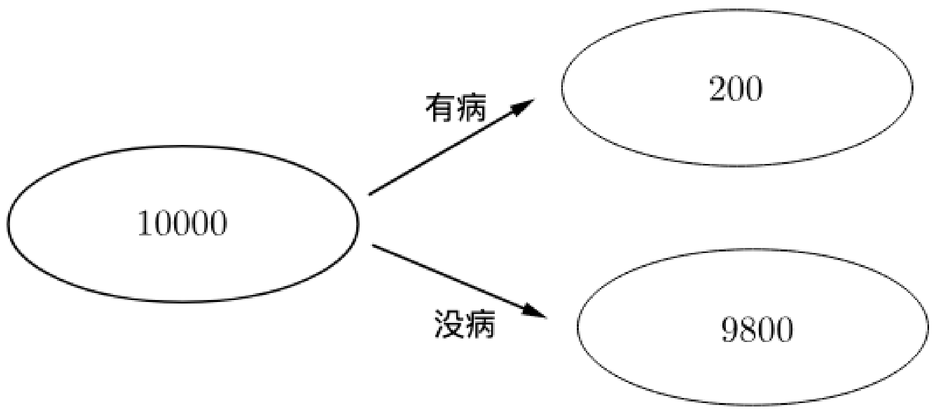
> 根据乘法公式可得
> $$P(A_1A_2B_3)=P(A_1)P(A_2|A_1)P(B_3|A_1A_2)$$
> $$=\frac{1}{2}\times\frac{6}{10}\times\frac{4}{10+2}=\frac{1}{10}$$
> 同样的类似方法可得
> $$P(A_1B_2A_3)=\frac{4}{8}\times\frac{4}{8+2}\times\frac{6}{10+2}=\frac{1}{10}$$
> $$P(B_1A_2A_3)=\frac{1}{2}\times\frac{4}{8+2}\times\frac{4+2}{10+2}=\frac{1}{10}$$
> 所以$P_2=\frac{3}{10}$

## 1.3 贝叶斯与全概率

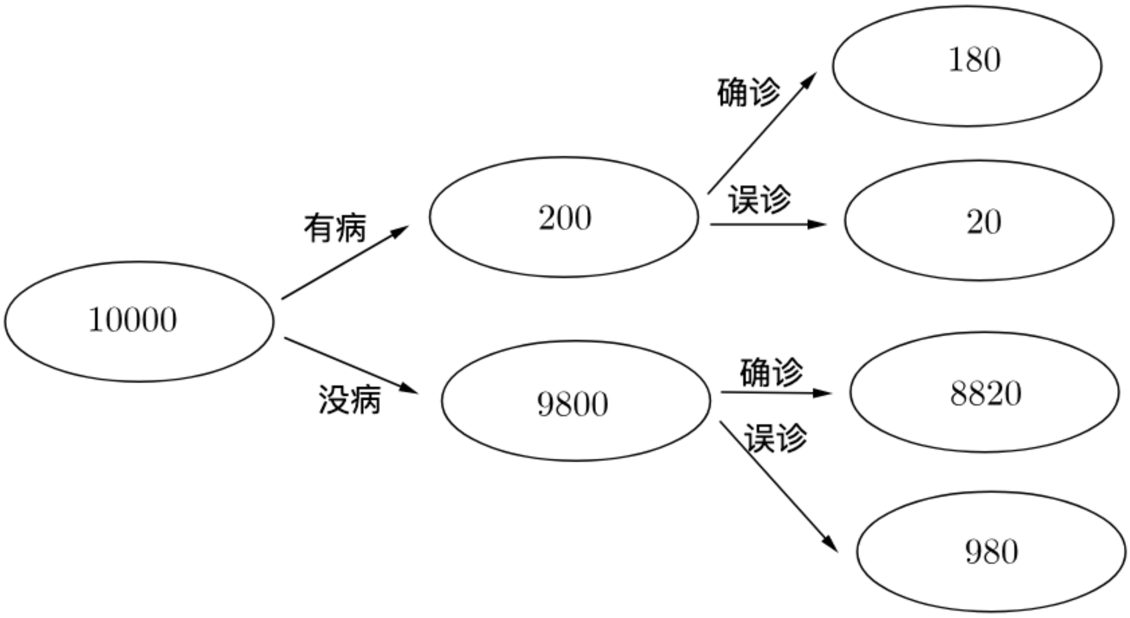
### 1.3.1 基本比例谬误

已知某种疾病在人群中有$2\%$的患病率，可以通过某种手段来检测这种疾病，这种手段有$90\%$的正确率。$90\%$的正确率很高嘛，似乎被诊断出有病了之后那基本就真的有病了。真的是这样吗？

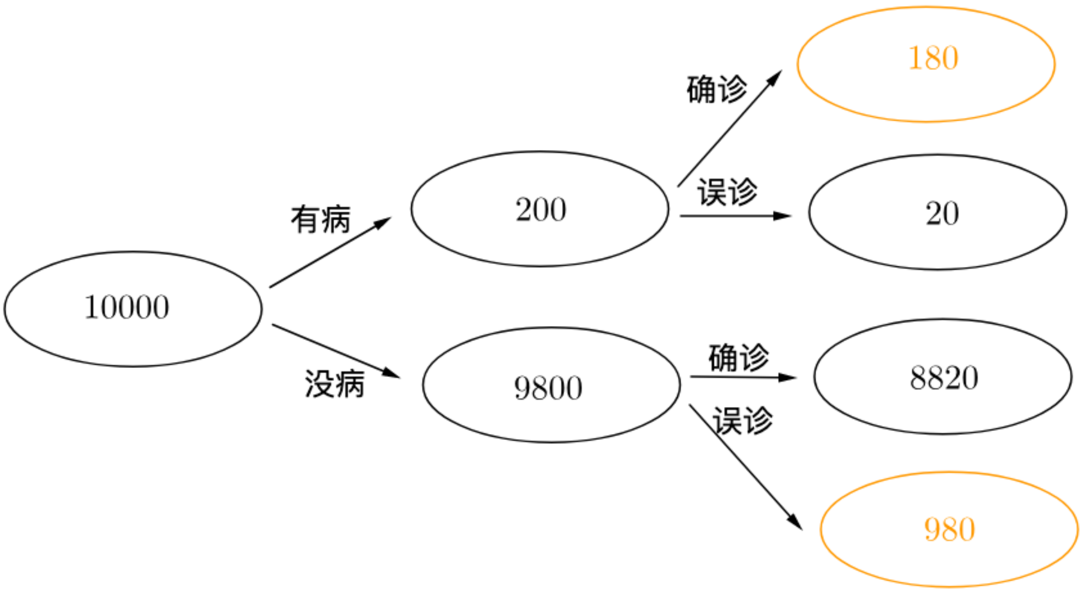
还是来画下树形图，假设有$10000$个人，$2\%$的患病率就是$200$个人有病：




正确率$90\%$，这个正确指的是，检测显示有病的是真有病，检测显示没病的是真没病（下图中“确诊”表示正确的诊断，“误诊”表示错误的诊断）：



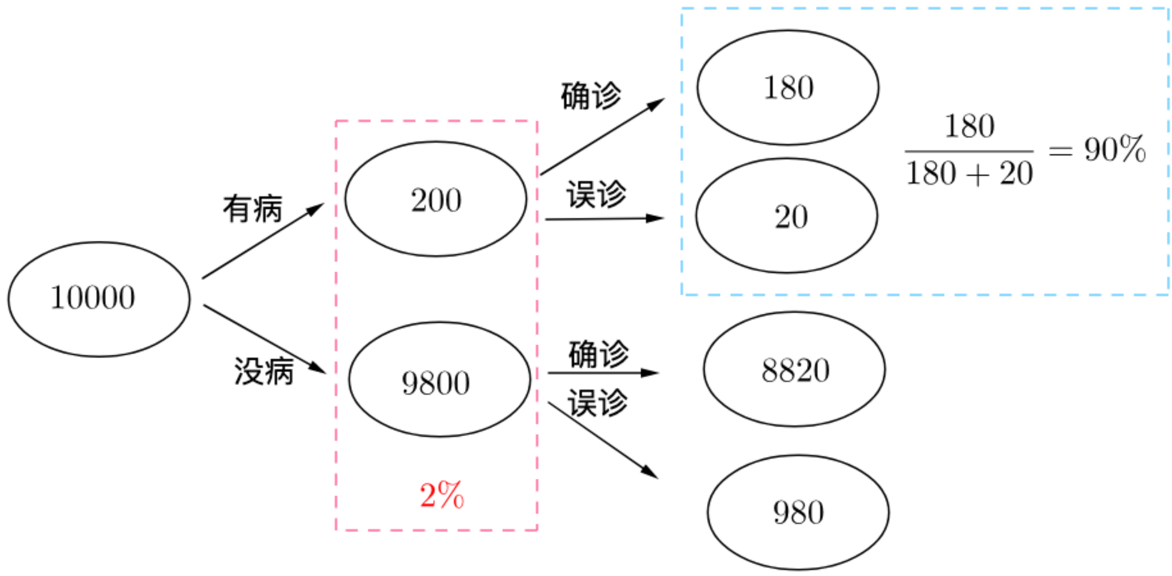

被诊断出有病包含两种可能性，一种是有病人群中的"确诊"和没病人群中的"误诊"：




也就是说这$10000$人里面会有$180+980=1160$人被诊断有病，而真正有病的不过$180$人，所以如果被检测出有病，真正有病的概率只有：

$$\frac{180}{1160} \approx 15.5\%$$

根本的原因在于忽视了人群分布的基本比例，虽然确诊率高达$90\%$，但是真正患病的人实在太少了，只占所有人中的$2\%$：




没患病的足足有$98\%$，哪怕误诊率为$10\%$，但人多啊，依然有很多被误诊为有病。

这种忽视基本比例导致的错误称为基本比例谬误（Base Rate Fallacy）。

## 1.3.2 贝叶斯

### 1.3.2.1 贝叶斯定理

对于随机事件$A$、$B$，若$P(B)\neq 0$，有：

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

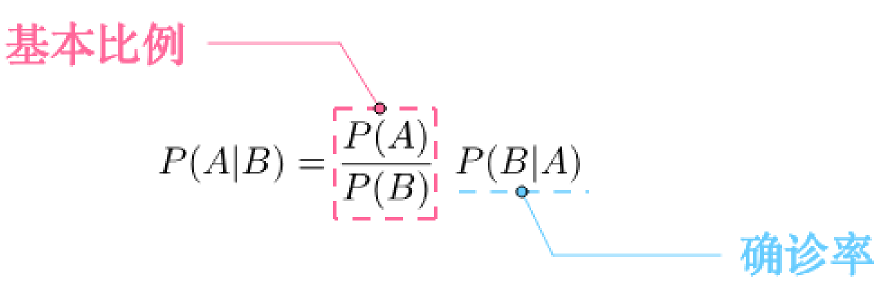
**解读1**： 在诊断问题中，已知检测的正确率有$90\%$，或者说确诊率为$90\%$：

$$P(B|A)=90\%$$

而求出来的"被诊断出有病了之后真的有病的概率"为：

$$P(A|B)\approx 15.5\%$$

之前说过，差别这么大在于基本比例悬殊，也就是所有人中患者只占了$2\%$，贝叶斯定理正好反映了这点：



$P(A)$、$P(B)$是关于所有人的概率：

$$P(A) = \text{所有人患病的概率}, \quad P(B) = \text{所有人检测显示有病的概率}$$

两者的比例就是所谓的基本比例。

**解读2**： 贝叶斯定理除了用于解读基本比例谬误以外，还有一种更常用的解释。先把贝叶斯定理变下形：

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

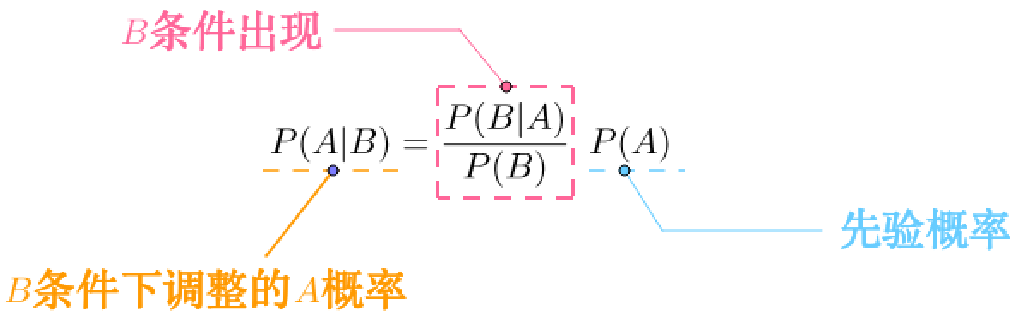
还是之前的疾病检测问题，还是之前定义的两个事件：

$$A = \text{“患病的人”}, \quad B = \text{“检测显示有病的人”}$$

一开始就知道人群中有$2\%$的患病率，这个也称为先验概率（大概就是最先就知道的意思），即：

$$P(A) = 2\%$$

如果只知道这个信息，那么你患病的概率也是$2\%$。但有一天去体检，检查显示你有病，这个时候相当于出现了新的信息，或者说$B$条件出现了，可以通过贝叶斯定理对患病的概率进行调整：




调整的结果当然是真正有病的概率急剧上升：

$$\begin{aligned} P(A|B) &= \frac{P(B|A)}{P(B)}P(A) \\ &= \frac{0.9}{0.02 \cdot 0.9 + 0.98 \cdot 0.1} \times 0.02 \\ &\approx 15.5\% \end{aligned}$$

> 例题
>
> $A$市有两种颜色的出租车，颜色和比例如下：


某天有一辆出租车夜间肇事后逃逸，但还好当时有一位目击证人，这位目击者认定肇事的出租车是绿色的。但是，他“目击的可信度”如何呢？公安人员经过在相同环境下对该目击者进行“蓝绿”测试而得到：$80\%$的情况下识别正确，$20\%$的情况不正确。

那么肇事车辆为绿色的概率有多大？

假设事件为：

$$A = \text{“绿色的出租车”}, \quad B = \text{“目击为绿色”}$$

在未知的情况下，全市的出租车都有可能肇事逃逸。那么绿色出租车肇事的概率和绿色出租车的比例是一样的，也就是先验概率为：

$$P(A) = 0.1$$

这个时候目击证人出现了，他说看到的是绿色的出租车，但是警方检测之后发现他的正确率是$80\%$，也就是：

$$P(B|A) = 0.8$$

被他目击为绿色的包括绿色出租车中的$80\%$、以及蓝色出租车中的$20\%$，即：

$$P(B) = 0.1 \cdot 0.8 + 0.9 \cdot 0.2$$

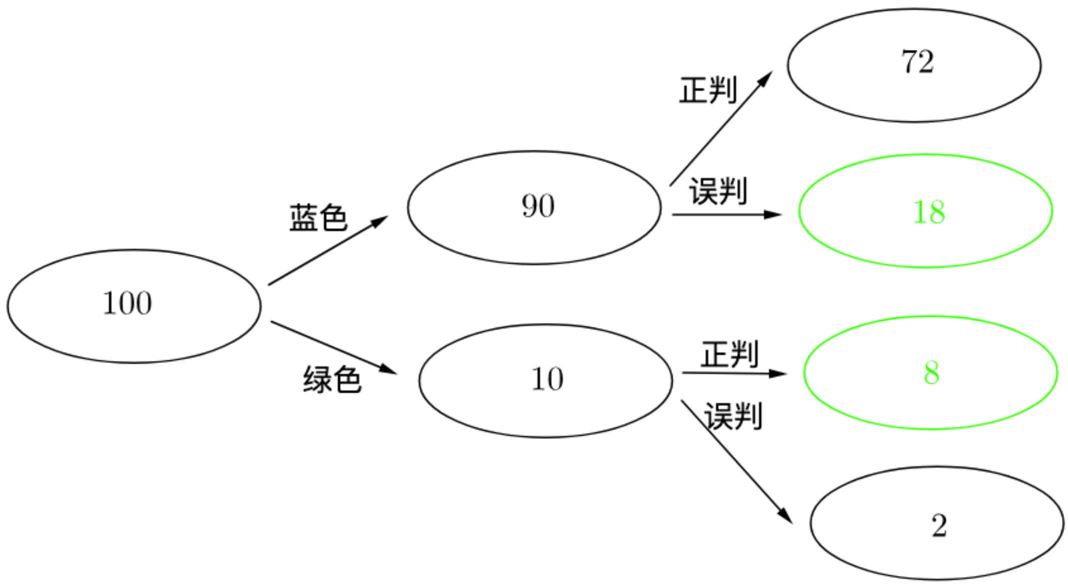
既然目击证人有了，那么肇事车辆为绿色的概率通过贝叶斯调整后，自然会变高：

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

$$= \frac{0.8}{0.1 \cdot 0.8 + 0.9 \cdot 0.2} \times 0.1$$

$$= \frac{4}{13}$$

当然你也可以用树形图来解，答案也是一样的：



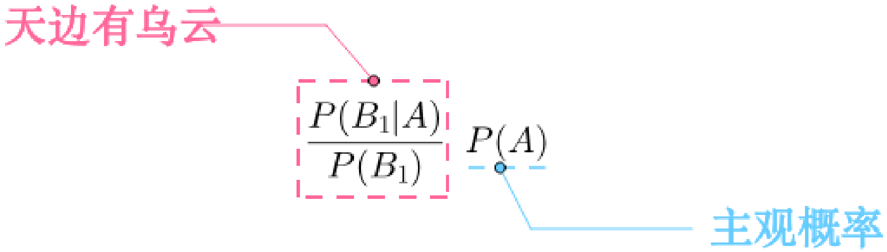

> 可以看出，哪怕有目击证人的情况下，绿色出租车作案的可能性仍然不高，因为基础比例在那里起作用，所以警察办案不能靠孤立的证据，很容易冤枉好人。

#### 1.3.2.1 贝叶斯与修正

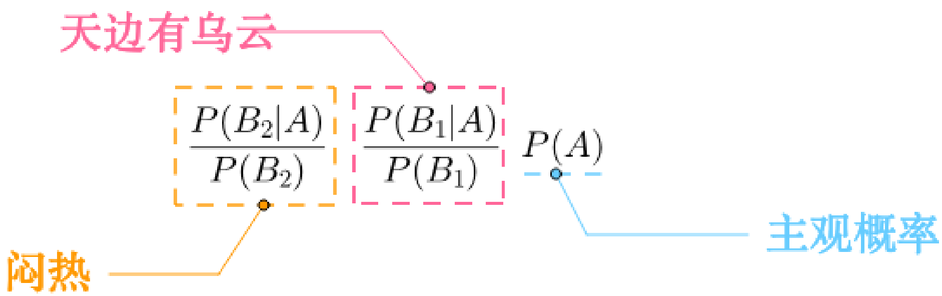
贝叶斯定理这种会被新的信息修正的特性特别受主观派的青睐。举个例子，比如要判断明天是否下雨，主观派说，反正一无所知就假设为$50\%$吧：

$$P(A)=50\%$$

突然发现，天边似乎有乌云，假设为事件$B_1$，那用贝叶斯修正一下：




又觉得天气闷热的，假设为事件$B_2$，再用贝叶斯修正一下：



通过不断地信息修正，主观概率会越来越变得客观（对此依然有很多争论，我们无意讨论）：

$$\text{客观概率} = \boxed{\frac{P(B_n|A)}{P(B_n)}} \cdots \boxed{\frac{P(B_2|A)}{P(B_2)}} \boxed{\frac{P(B_1|A)}{P(B_1)}} \underbrace{P(A)}_{\text{主观概率}}$$