Conditional Probability, Independence, and Combinations of Events

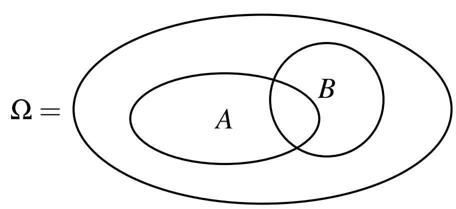
One of the key properties of coin flips is independence: if you flip a fair coin ten times and get ten T's, this does not make it more likely that the next coin flip will be H's. It still has exactly 50% chance of being H.

By contrast, suppose while dealing cards, the first ten cards are all red (hearts or diamonds). What is the chance that the next card is red? We started with exactly 26 red cards and 26 black cards. But after dealing the first ten cards, we know that the deck has 16 red cards and 26 black cards. So the chance that the next card is red is $\frac{16}{42}$. So unlike the case of coin flips, now the chance of drawing a red card is no longer independent of the previous card that was dealt. This is the phenomenon we will explore in this note on conditional probability.

1 Conditional Probability

Let's consider an example with a smaller sample space. Suppose we toss $m = 4$ labeled balls into $n = 3$ labeled bins; this is a uniform sample space with $3^4 = 81$ sample points. From the previous note, we already know that the probability the first bin is empty is $(1-\frac{1}{3})^4 = (\frac{2}{3})^4 = \frac{16}{81}$. What is the probability of this event *given* that the second bin is empty? Let A denote the event that the first bin is empty, and B the event that the second bin is empty. In the language of conditional probability, we wish to compute the probability $\mathbb{P}[A|B]$, which we read as "the conditional probability of A given B."

How should we compute $\mathbb{P}[A|B]$? Since event B is guaranteed to happen, we need to look not at the whole sample space Ω, but at the smaller sample space consisting only of the sample points in B. In terms of the picture below, we are no longer looking at the large oval, but only the oval labeled B:



What should be the probability of each sample point $\omega \in B$ given that the event B occurs? If they all simply inherited their probabilities from Ω, then the sum of these probabilities would be $\sum_{\omega\in B}\mathbb{P}[\omega] = \mathbb{P}[B]$, which in general is less than 1. So, to get the correct normalization, we need to *scale* the probability of each sample point by $\frac{1}{\mathbb{P}[B]}$. That is, for each sample point $\omega \in B$, the new probability becomes

$$\mathbb{P}[\omega|B] = \frac{\mathbb{P}[\omega]}{\mathbb{P}[B]}.$$

Now it is clear how to compute $\mathbb{P}[A|B]$: namely, we just sum up these scaled probabilities over all sample

EECS 70 离散数学与概率论 2020年秋季

笔记14

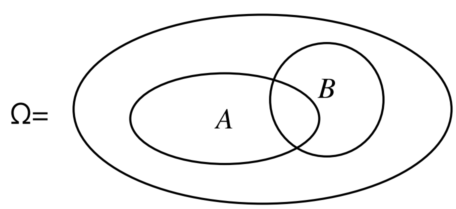
条件概率、独立性以及事件的组合

抛硬币的一个关键性质是独立性：如果你连续十次抛一枚公平硬币都得到反面（T），这并不会使下一次抛出正面（H）的可能性增加。下一次仍然是恰好50%的概率出现H。相比之下，假设在发牌过程中，前十个牌都是红色（红心或方块）。那么下一张牌是红色的概率是多少？我们一开始正好有26张红牌和26张黑牌。但在发出前十张牌后，我们知道牌堆中剩下16张红牌和26张黑牌。因此下一张牌是红色的概率为16/42。因此，与抛硬币的情况不同，现在抽到红牌的概率不再独立于之前发出的牌。这就是本笔记中我们将探讨的条件概率现象。

1 条件概率

让我们考虑一个样本空间较小的例子。假设我们将m = 4个带标签的球投入n = 3个带标签的箱子中；这是一个均匀样本空间，共有3^4 = 81个样本点。从前一篇笔记中我们已知第一个箱子为空的概率是$(1-1/3)^4 = (2/3)^4 = 16/81$。在已知第二个箱子为空的前提下，第一个箱子为空的概率是多少？设A表示第一个箱子为空的事件，B表示第二个箱子为空的事件。用条件概率的语言，我们希望计算P[A|B]，即在给定B条件下A的条件概率。

我们应如何计算P[A|B]？由于事件B必然发生，我们需要关注的不再是整个样本空间Ω，而是仅包含B中样本点的更小样本空间。用下图的语言来说，我们不再看大的椭圆，而只看标记为B的椭圆：



当事件B发生时，每个样本点ω∈B的概率应如何设定？如果它们直接继承自Ω中的概率，那么这些概率之和将是Σω∈B P[ω] = P[B]，通常小于1。因此，为了正确归一化，我们需要将每个样本点的概率按$\frac{1}{\mathbb{P}[B]}$。也就是说，对于每个样本点ω∈B，新的概率变为

$$\mathbb{P}[\omega|B] = \frac{\mathbb{P}[\omega]}{\mathbb{P}[B]}.$$

现在很清楚如何计算P[A|B]：我们只需将这些缩放后的概率在所有样本点上求和

points that lie in both A and B:

$$\mathbb{P}[A|B] = \sum_{\omega \in A\cap B} \mathbb{P}[\omega|B] = \sum_{\omega \in A\cap B} \frac{\mathbb{P}[\omega]}{\mathbb{P}[B]} = \frac{\mathbb{P}[A\cap B]}{\mathbb{P}[B]}.$$

Definition 14.1 (Conditional Probability)**.** *For events $A, B \subseteq \Omega$ in the same probability space such that $\mathbb{P}[B] > 0$, the conditional probability of A given B is*

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A\cap B]}{\mathbb{P}[B]}.$$

Returning to our example of balls and bins, to compute $\mathbb{P}[A|B]$ we need to figure out $\mathbb{P}[A\cap B]$. But $A\cap B$ is the event that both the first two bins are empty, i.e., all four balls fall in the third bin. So $\mathbb{P}[A\cap B] = \frac{1}{81}$ (why?). Therefore,

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A\cap B]}{\mathbb{P}[B]} = \frac{1/81}{16/81} = \frac{1}{16}.$$

Not surprisingly, $\mathbb{P}[A|B] = \frac{1}{16} = 0.0625$ is quite a bit less than $\mathbb{P}[A] = \frac{16}{81} \approx 0.1975$; knowing that bin 2 is empty makes it significantly less likely that bin 1 will be empty.

Example: Card Dealing

Let's apply the ideas discussed above to compute the probability that, when dealing 2 cards and the first card is known to be an ace, the second card is also an ace. Let B be the event that the first card is an ace, and let A be the event that the second card is an ace.

To compute $\mathbb{P}[A|B]$, we need to figure out $\mathbb{P}[A\cap B]$. This is the probability that both cards are aces. Note that there are $52\cdot 51$ sample points in the sample space, since each sample point is a sequence of two cards. A sample point is in $A\cap B$ if both cards are aces. This can happen in $4\cdot 3 = 12$ ways.

Since each sample point is equally likely, $\mathbb{P}[A\cap B] = \frac{12}{52\cdot 51}$, while $\mathbb{P}[B]$, the probability of drawing an ace in the first trial, is $\frac{4}{52}$. Therefore,

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A\cap B]}{\mathbb{P}[B]} = \frac{3}{51},$$

which is less than $\mathbb{P}[A] = \frac{4}{52} = \frac{1}{13}$ (see Exercise 1). Hence, if the first card is an ace, it makes it less likely that the second card is also an ace.

2 Bayesian Inference

Now that we have introduced the notion of conditional probability, we can see how it is used in real world settings. Conditional probability is at the heart of a subject called *Bayesian inference*, used extensively in fields such as machine learning, communications and signal processing. Bayesian inference is a way to *update knowledge* after making an observation. For example, we may have an estimate of the probability of a given event A. After event B occurs, we can update this estimate to $\mathbb{P}[A|B]$. In this interpretation, $\mathbb{P}[A]$ can be thought of as a *prior* probability: our assessment of the likelihood of an event of interest, A, *before* making an observation. It reflects our prior knowledge. $\mathbb{P}[A|B]$ can be interpreted as the *posterior* probability of A after the observation. It reflects our updated knowledge.

Here is an example of where we can apply such a technique. A pharmaceutical company is marketing a new test for a certain medical disorder. According to clinical trials, the test has the following properties:

同时位于A和B中的样本点：

$$\mathbb{P}[A|B] = \sum_{\omega\in A\cap B} \mathbb{P}[\omega|B] = \sum_{\omega\in A\cap B} \frac{\mathbb{P}[\omega]}{\mathbb{P}[B]} = \frac{\mathbb{P}[A\cap B]}{\mathbb{P}[B]}.$$

定义14.1（条件概率）。对于同一概率空间中满足P[B] > 0的事件A、B ⊆ Ω，A在给定B下的条件概率为

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A\cap B]}{\mathbb{P}[B]}.$$

回到我们关于球与箱子的例子，要计算P[A|B]，我们需要确定P[A∩B]。但A∩B表示前两个箱子都为空的事件，即所有四个球都落入第三个箱子。因此P[A∩B] = $\frac{1}{81}$（为什么？）。因此，

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A\cap B]}{\mathbb{P}[B]} = \frac{1/81}{16/81} = \frac{1}{16}.$$

不出所料，P[A|B] = $\frac{1}{16}$ = 0.0625 远小于 P[A] = $\frac{16}{81}$ ≈0.1975；知道箱子2为空知道第二个箱子为空会使第一个箱子为空的可能性显著降低。

示例：发牌

让我们应用上述讨论的思想来计算发两张牌时，在已知第一张是A的情况下，第二张也是A的概率。设B为第一张是A的事件，A为第二张是A的事件。

要计算P[A|B]，我们需要确定P[A∩B]。这是两张牌都是A的概率。注意样本空间中有52·51个样本点，因为每个样本点是由两张牌组成的序列。当两张牌都是A时，该样本点属于A∩B。这种情况有4·3 = 12种方式发生。

由于每个样本点等可能发生，P[A∩B] = $\frac{12}{52\cdot 51}$而P[B]表示在第一次抽牌时抽到A的概率，第一次抽牌时的概率是452。因此，

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A\cap B]}{\mathbb{P}[B]} = \frac{3}{51},$$

这小于 P[A] = $\frac{4}{52}$ = $\frac{1}{13}$（见练习1）。因此，如果第一张牌是A，则第二张牌也是A的可能性会降低第二张牌也是A的可能性。

2 贝叶斯推断

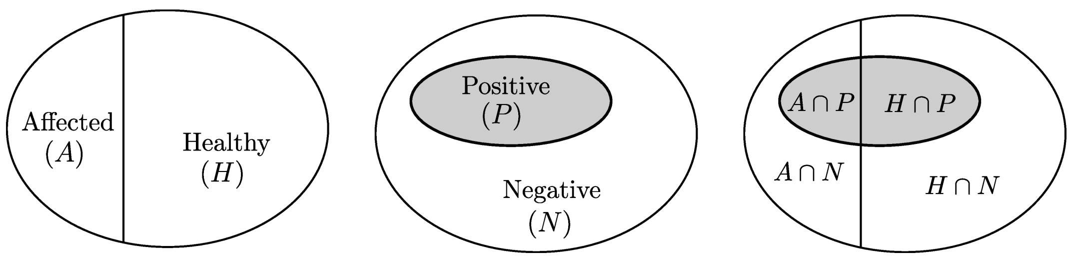
现在我们已经引入了条件概率的概念，可以看到它在现实世界中的应用。条件概率是贝叶斯推断的核心，广泛应用于机器学习、通信和信号处理等领域。贝叶斯推断是一种在进行观测后更新知识的方法。例如，我们可能对某个事件A发生的概率有一个估计值。在事件B发生之后，我们可以将其更新为P[A|B]。在这种解释中，P[A]可以被视为先验概率：即在进行观测之前，我们对感兴趣事件A发生可能性的评估，反映了我们的先验知识。P[A|B]则可被解释为在观测之后事件A的后验概率，反映了我们更新后的知识。

以下是一个我们可以应用这种技术的例子。一家制药公司正在推广一种针对某种医学疾病的新型检测方法。根据临床试验，该测试具有以下特性：

1. When applied to an affected person, the test comes up positive in 90% of cases, and negative in 10% (these are called "false negatives").

2. When applied to a healthy person, the test comes up negative in 80% of cases, and positive in 20% (these are called "false positives").

Suppose that the incidence of the disorder in the US population is 5%; this is our prior knowledge. When a random person is tested and the test comes up positive, how can we update the probability that the person has the disorder? (Note that this is presumably *not* the same as the simple probability that a random person in the population has the disorder, which is just $\frac{1}{20}$.) The implicit probability space here is the entire US population with equal probabilities.



The sample space here consists of all people in the US — denote their number by N (so $N \approx 325$ million). Let A be the event that a person chosen at random is affected, and B be the event that a person chosen at random tests positive. Now we can rewrite the information above as:

- $\mathbb{P}[A] = 0.05$, (5% of the U.S. population is affected)
- $\mathbb{P}[B|A] = 0.9$, (90% of the affected people test positive)
- $\mathbb{P}[B|\overline{A}] = 0.2$, (20% of healthy people test positive)

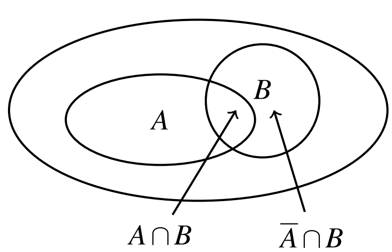
We want to calculate $\mathbb{P}[A|B]$. We can proceed as follows:

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} = \frac{\mathbb{P}[B|A]\mathbb{P}[A]}{\mathbb{P}[B]}. \tag{1}$$

We obtained the second equality above by applying the definition of conditional probability:

$$\mathbb{P}[B|A] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[A]}.$$

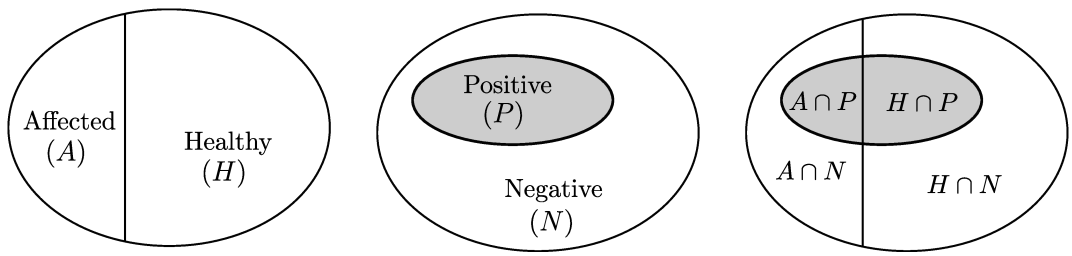
To evaluate (1), we need to compute $\mathbb{P}[B]$. This is the probability that a random person tests positive. To compute this, we can sum two values: the probability $\mathbb{P}[\overline{A} \cap B]$ that a healthy person tests positive and the probability $\mathbb{P}[A \cap B]$ that an affected person tests positive. We can sum because the events $\overline{A} \cap B$ and $A \cap B$ do not intersect:



1. 对患者使用该测试时，90% 的情况下结果为阳性，10% 为阴性（这些被称为假阴性）。

2. 对健康人使用该测试时，80% 的情况下结果为阴性，20% 为阳性（这些被称为假阳性）。

假设该疾病在美国人群中的发病率为 5%；这是我们先验的知识。当对一个随机人选进行测试且结果呈阳性时，我们如何更新此人患有该疾病的概率？（注意，这大概不等于人群中随机一人患病的简单概率，后者仅为 $\frac{1}{20}$）此处隐含的概率空间是整个美国人群中以相等概率选择。



这里的样本空间包括所有美国人——用 N 表示其人数（因此 N ≈3.25 亿）。设 A 为随机选择的人患病的事件，B 为随机选择的人测试呈阳性的事件。现在我们可以将上述信息重写为：

- P[A] = 0.05，（美国人口的 5% 受影响）
- P[B|A] = 0.9，（90% 的患者测试呈阳性）
- P[B|$\overline{\text{A}}$] = 0.2，（20% 的健康人测试呈阳性）我

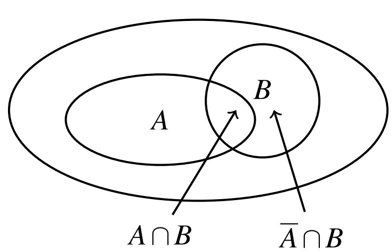
们想要计算 P[A|B]。我们可以按如下步骤进行：

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} = \frac{\mathbb{P}[B|A]\mathbb{P}[A]}{\mathbb{P}[B]}. \tag{1}$$

我们通过对条件概率定义的应用得到了上面的第二个等式：

$$\mathbb{P}[B|A] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[A]}.$$

为了计算 (1)，我们需要求出 P[B]。这是指一个人随机测试呈阳性的概率。为此，我们可以将两个值相加：健康人测试呈阳性的概率 P[A∩B]，以及患病者测试呈阳性的概率 P[A∩B]。我们可以相加是因为事件 A∩B 和 A∩B 不相交：



By again applying the definition of conditional probability, we have:

$$\begin{aligned}\mathbb{P}[B] = \mathbb{P}[A\cap B]+\mathbb{P}[\overline{A}\cap B] &= \mathbb{P}[B|A]\mathbb{P}[A]+\mathbb{P}[B|\overline{A}]\mathbb{P}[\overline{A}] \\ &= \mathbb{P}[B|A]\mathbb{P}[A]+\mathbb{P}[B|\overline{A}](1-\mathbb{P}[A]).\end{aligned} \tag{2}$$

Combining (1) and (2), we have expressed $\mathbb{P}[A|B]$ in terms of $\mathbb{P}[A],\mathbb{P}[B|A]$ and $\mathbb{P}[B|\overline{A}]$:

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[B|A]\mathbb{P}[A]}{\mathbb{P}[B|A]\mathbb{P}[A]+\mathbb{P}[B|\overline{A}](1-\mathbb{P}[A])}. \tag{3}$$

By plugging in the values written above, we obtain $\mathbb{P}[A|B]=\frac{9}{47}\approx 0.19$.

Equation (3) is useful for many inference problems. We are given $\mathbb{P}[A]$, which is the (unconditional) probability that the event of interest, A, happens. We are given $\mathbb{P}[B|A]$ and $\mathbb{P}[B|\overline{A}]$, which quantify how noisy the observation is. (If $\mathbb{P}[B|A]=1$ and $\mathbb{P}[B|\overline{A}]=0$, for example, the observation is completely noiseless.) Now we want to calculate $\mathbb{P}[A|B]$, the probability that the event of interest happens given we made the observation. Equation (3) allows us to do just that.

Of course, (1), (2) and (3) are derived from the basic axioms of probability and the definition of conditional probability, and are therefore true with or without the above Bayesian inference interpretation. However, this interpretation is very useful when we apply probability theory to study inference problems.

3 Bayes' Rule and Total Probability Rule

Equations (1) and (2) are very useful in their own right. The former is called **Bayes' Rule** and the latter is called the **Total Probability Rule**. Bayes' Rule is useful when one wants to calculate $\mathbb{P}[A|B]$ but one is given $\mathbb{P}[B|A]$ instead, i.e., it allows us to "flip" things around.

The Total Probability Rule is an application of the strategy of "dividing into cases." There are two possibilities: either an event A happens or A does not happen. If A happens, the probability that B happens is $\mathbb{P}[B|A]$. If A does not happen, the probability that B happens is $\mathbb{P}[B|\overline{A}]$. If we know or can easily calculate these two probabilities and also $\mathbb{P}[A]$, then the Total Probability Rule yields the probability of event B.

3.1 Examples

Tennis Match

You are about to play a tennis match against a randomly chosen opponent and you wish to calculate your probability of winning. You know your opponent will be one of two people, X or Y. If person X is chosen, you will win with probability 0.7. If person Y is chosen, you will win with probability 0.3. Your opponent is chosen by flipping a biased coin such that the probability of choosing X is 0.6.

Let's first determine which events we are interested in. Let A be the event that you win. Let B_X be the event that person X is chosen, and let B_Y be the event that person Y is chosen. We wish to calculate $\mathbb{P}[A]$. Here is what we know so far:

- $\mathbb{P}[A|B_X]=0.7$, (if person X is chosen, you win with probability 0.7)
- $\mathbb{P}[A|B_Y]=0.3$, (if person Y is chosen, you win with probability 0.3)

再次应用条件概率的定义，我们有：

$$\mathbb{P}[B]=\mathbb{P}[A\cap B]+\mathbb{P}[\overline{A}\cap B]=\mathbb{P}[B|A]\mathbb{P}[A]+\mathbb{P}[B|\overline{A}]\mathbb{P}[\overline{A}] \tag{2}$$

$$=\mathrm{P}[B|A]\mathrm{P}[A]+\mathrm{P}[B|\overline{A}](1-\mathrm{P}[A])。$$

结合 (1) 和 (2)，我们已将 P[A|B] 表示为 P[A]、P[B|A] 和 P[B|$\overline{\text{A}}$] 的函数：

$$\mathbb{P}[A|B]=\frac{\mathbb{P}[B|A]\mathbb{P}[A]}{\mathbb{P}[B|A]\mathbb{P}[A]+\mathbb{P}[B|\overline{A}](1-\mathbb{P}[A])}. \tag{3}$$

代入上面写出的数值，我们得到 P[A|B] = $\frac{9}{47}$ ≈0.19。

方程 (3) 对许多推断问题都很有用。我们已知 P[A]，即感兴趣事件 A 发生的（无条件）概率。我们已知 P[B|A] 和 P[B|$\overline{\text{A}}$]，它们量化了观测的噪声程度。（例如，若 P[B|A] = 1 且 P[B|$\overline{\text{A}}$] = 0，则观测是完全无噪声的。）现在我们希望计算 P[A|B]，即在进行了观测后，感兴趣事件发生的概率。方程 (3) 使我们能够做到这一点。

当然，(1)、(2) 和 (3) 都是从概率的基本公理和条件概率的定义推导而来的，因此无论是否有上述贝叶斯推断解释，它们都成立。然而，这种解释在我们将概率论应用于研究推断问题时非常有用。

3 贝叶斯法则与全概率法则

方程 (1) 和 (2) 本身非常有用。前者称为贝叶斯法则，后者称为全概率法则。当我们想计算 P[A|B] 但给定的是 P[B|A] 时，贝叶斯法则是有用的，也就是说，它允许我们“反过来”计算。

全概率公式是分情况讨论策略的一种应用。有两种可能性：事件 A 发生或 A 不发生。若 A 发生，则 B 发生的概率为 P[B|A]；若 A 不发生，则 B 发生的概率为 P[B|$\overline{\text{A}}$]。如果我们知道或可以轻易计算这两个概率以及 P[A]，那么全概率公式就能给出事件 B 的概率。

3.1 示例

网球比赛

你即将与一位随机选定的对手进行一场网球比赛，并希望计算你的获胜概率。你知道你的对手将是两个人之一，X 或 Y。如果选择了 X，则你获胜的概率为 0.7；如果选择了 Y，则你获胜的概率为 0.3。对手通过抛一枚有偏硬币来决定，选择 X 的概率为 0.6。

我们首先确定所关心的事件。设 A 为你获胜的事件。设 BX 为人 X 被选中的事件，BY 为人 Y 被选中的事件。我们希望计算 P[A]。以下是我们目前所知道的信息：

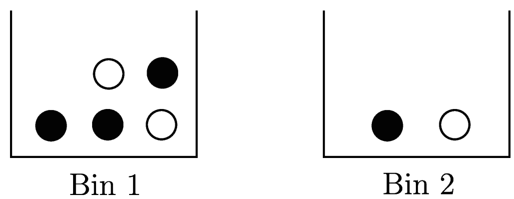
- P[A|BX] = 0.7，（如果选择了人 X，则你获胜的概率为 0.7）
- P[A|BY] = 0.3，（如果选择了人 Y，则你获胜的概率为 0.3）

- $\mathbb{P}[B_X] = 0.6$, (person X is chosen with probability 0.6)
- $\mathbb{P}[B_Y] = 0.4$, (person Y is chosen with probability 0.4)

By using the Total Probability Rule, we have $\mathbb{P}[A] = \mathbb{P}[A|B_X]\mathbb{P}[B_X] + \mathbb{P}[A|B_Y]\mathbb{P}[B_Y]$, and plugging in the known values gives $\mathbb{P}[A] = (0.7 \times 0.6) + (0.3 \times 0.4) = 0.54$.

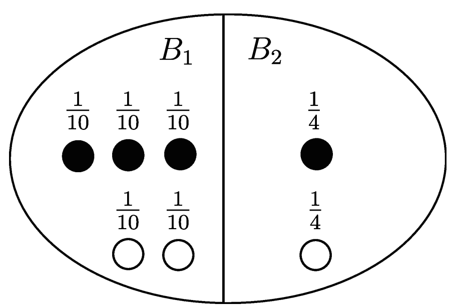
Balls and Bins

Imagine we have the following two bins each containing some number of black and white balls:



Suppose one of the two bins is chosen with equal probability and a ball is drawn from the chosen bin uniformly at random. What is the probability that we picked Bin 1 given that a white ball was drawn, i.e., $\mathbb{P}[\text{Bin } 1 \mid \bigcirc]$?

Is the answer $\frac{2}{3}$, since we know that there are a total of three white balls, two of which are in Bin 1? This reasoning is *incorrect*. Instead, what we should do is appropriately scale each sample point as the following picture shows:



This diagram shows that the sample space Ω consists of the outcomes in event B_1 (corresponding to Bin 1) and event B_2 (corresponding to Bin 2), i.e., $\Omega = B_1 \cup B_2$. We can use the definition of conditional probability to see that

$$\mathbb{P}[\text{Bin1} \mid \bigcirc] = \frac{\frac{1}{10} + \frac{1}{10}}{\frac{1}{10} + \frac{1}{10} + \frac{1}{4}} = \frac{\frac{2}{10}}{\frac{9}{20}} = \frac{4}{9}.$$

Let us try to achieve this probability using Bayes' Rule. To apply Bayes' Rule, we need to compute $\mathbb{P}[\bigcirc|\text{Bin 1}]$, $\mathbb{P}[\text{Bin 1}]$, and $\mathbb{P}[\bigcirc]$. Here, $\mathbb{P}[\bigcirc|\text{Bin 1}]$ is the chance that we pick a white ball given that we picked Bin 1, which is $\frac{2}{5}$. $\mathbb{P}[\text{Bin 1}] = \frac{1}{2}$, as given in the description of the problem. Finally, $\mathbb{P}[\bigcirc]$ can be computed using the Total Probability Rule:

$$\mathbb{P}[\bigcirc] = \mathbb{P}[\bigcirc|\text{Bin 1}] \times \mathbb{P}[\text{Bin 1}] + \mathbb{P}[\bigcirc|\text{Bin 2}] \times \mathbb{P}[\text{Bin 2}] = \frac{2}{5} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{9}{20}.$$

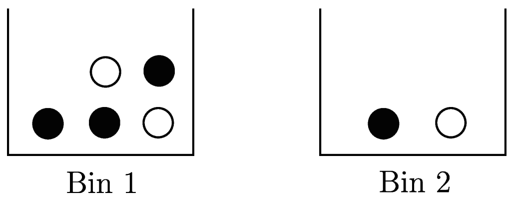
Observe that we can apply the Total Probability Rule here because $\mathbb{P}[\text{Bin 1}]$ is the complement of $\mathbb{P}[\text{Bin 2}]$. Finally, upon plugging the above values into Bayes' Rule, we obtain the probability that we picked Bin 1

- P[BX] = 0.6，（选择人 X 的概率为 0.6）
- P[BY] = 0.4，（选择人 Y 的概率为 0.4）

利用全概率公式，我们有 P[A] = P[A|BX]P[BX] + P[A|BY]P[BY]，代入已知值得到 P[A] = (0.7×0.6)+(0.3×0.4) = 0.54。

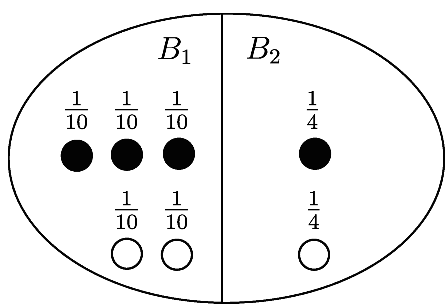
球与箱子

设想我们有两个箱子，每个箱子包含若干黑球和白球：



假设以相等的概率选择两个箱子中的一个，并从选中的箱子中均匀随机抽取一个球。已知抽到了一个白球，求我们选中的是箱子 1 的概率，即 P[箱子 1 | ❢]?

答案是 2 吗 3，因为我们知道总共有三个白球，其中两个在箱子 1 中？这个
推理是错误的。相反，我们应该像下图所示那样适当地对每个样本点进行缩放：



该图显示样本空间 Ω 由事件 B1（对应箱子 1）和事件 B2（对应箱子 2）的结果组成，即 Ω = B1∪B2。我们可以使用条件概率的定义来观察到

$$\mathbb{P}[\text{Bin1} \mid \bigcirc] = \frac{\frac{1}{10}+\frac{1}{10}}{\frac{1}{10}+\frac{1}{10}+\frac{1}{4}} = \frac{\frac{2}{10}}{\frac{9}{20}} = \frac{4}{9}.$$

让我们尝试使用贝叶斯法则来获得这个概率。要应用贝叶斯法则，我们需要计算 P[❢|Bin 1]、P[Bin 1] 和 P[❢]。这里，P[❢|Bin 1] 是在选择 Bin 1 的条件下抽到白球的概率，为 25。P[Bin 1] = 1，由问题描述给出。最后，P[❢] 可以通过全概率法则计算：

$$\mathbb{P}[\bigcirc] = \mathbb{P}[\bigcirc|\text{Bin 1}] \times \mathbb{P}[\text{Bin 1}] + \mathbb{P}[\bigcirc|\text{Bin 2}] \times \mathbb{P}[\text{Bin 2}] = \frac{2}{5} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{9}{20}.$$

注意到我们可以在此处应用全概率法则，因为 P[Bin 1] 是 P[Bin 2] 的补集。最后，将上述数值代入贝叶斯法则，我们得到抽到 Bin 1 的概率

given that we picked a white ball:

$$\mathbb{P}[\text{Bin } 1 \mid \bigcirc] = \frac{\frac{2}{5} \times \frac{1}{2}}{\frac{9}{20}} = \frac{\frac{2}{10}}{\frac{9}{20}} = \frac{4}{9}.$$

All we have done above is to combine Bayes' Rule and the Total Probability Rule; this is also how we obtained (3). Equivalently, we could have plugged in the appropriate values to (3).

3.2 Generalization

We now consider Bayes' Rule and the Total Probability Rule in a more general context. First, we define a *partition* of an event as follows.

Definition 14.2 (Partition of an event)**.** *We say that an event A is partitioned into n events $A_1,\dots,A_n$ if*

1. $A = A_1 \cup A_2 \cup \cdots \cup A_n$*, and*
2. $A_i \cap A_j = \varnothing$ *for all* $i \neq j$ *(i.e.,* $A_1,\dots,A_n$ *are mutually exclusive).*

In other words, each outcome in A belongs to exactly one of the events $A_1,\dots,A_n$.

Now, let $A_1,\dots,A_n$ be a partition of the sample space Ω. Then, the **Total Probability Rule** for any event B is

$$\mathbb{P}[B] = \sum_{i=1}^{n} \mathbb{P}[B \cap A_i] = \sum_{i=1}^{n} \mathbb{P}[B|A_i]\,\mathbb{P}[A_i], \tag{4}$$

while **Bayes' Rule**, assuming $\mathbb{P}[B] \neq 0$, is given by

$$\mathbb{P}[A_i|B] = \frac{\mathbb{P}[B|A_i]\,\mathbb{P}[A_i]}{\mathbb{P}[B]} = \frac{\mathbb{P}[B|A_i]\,\mathbb{P}[A_i]}{\sum_{j=1}^{n} \mathbb{P}[B|A_j]\,\mathbb{P}[A_j]}, \tag{5}$$

where the second equality follows from the Total Probability Rule.

4 Combinations of Events

In most applications of probability in Computer Science, we are interested in things like $\mathbb{P}[\bigcup_{i=1}^{n} A_i]$ and $\mathbb{P}[\bigcap_{i=1}^{n} A_i]$, where the A_i are simple events (i.e., we know or can easily compute $\mathbb{P}[A_i]$). The intersection $\bigcap_i A_i$ corresponds to the logical AND of the events A_i, while the union $\bigcup_i A_i$ corresponds to their logical OR. As an example, if A_i denotes the event that a failure of type i happens in a certain system, then $\bigcup_i A_i$ is the event that the system fails.

In general, computing the probabilities of such combinations can be very difficult. In this section, we discuss some situations where it can be done. Let's start with independent events, for which intersections are quite simple to compute.

4.1 Independent Events

Definition 14.3 (Independence)**.** *Two events A, B in the same probability space are said to be independent if* $\mathbb{P}[A \cap B] = \mathbb{P}[A] \times \mathbb{P}[B]$.

已知我们抽到了一个白球：

$$\mathbb{P}[\text{Bin } 1 \mid \bigcirc] = \frac{\frac{2}{5} \times \frac{1}{2}}{\frac{9}{20}} = \frac{\frac{2}{10}}{\frac{9}{20}} = \frac{4}{9}.$$

我们上面所做的只是将贝叶斯法则和全概率法则结合起来；这也是我们如何得到 (3) 的方式。等价地，我们可以将适当的值代入 (3)。

3.2 推广

我们现在在更一般的背景下考虑贝叶斯法则和全概率法则。首先，我们如下定义事件的划分。

定义 14.2 (事件的划分)。我们说一个事件 A 被划分为 n 个事件 A1,...,An，如果

1. A = A1∪A2∪···∪An，且
2. Ai ∩Aj = ∅ 对所有 i ≠ j 成立（即 A1,...,An 是互斥的）。

换句话说，A 中的每个结果恰好属于事件 A1,...,An 中的一个。

现在，设 A1,...,An 是样本空间 Ω 的一个划分。那么，对任意事件 B 的全概率法则为

$$\mathbb{P}[B] = \sum_{i=1}^{n} \mathbb{P}[B \cap A_i] = \sum_{i=1}^{n} \mathbb{P}[B|A_i]\,\mathbb{P}[A_i], \tag{4}$$

而贝叶斯法则，在假设 P[B] ≠ 0 的情况下，表示为

$$\mathbb{P}[A_i|B] = \frac{\mathbb{P}[B|A_i]\,\mathbb{P}[A_i]}{\mathbb{P}[B]} = \frac{\mathbb{P}[B|A_i]\,\mathbb{P}[A_i]}{\sum_{j=1}^{n} \mathbb{P}[B|A_j]\,\mathbb{P}[A_j]}, \tag{5}$$

其中第二个等式来源于全概率法则。

4 事件的组合

在计算机科学的概率应用中，我们通常关注的是 P[Sni=1Ai] 和 P[Tni=1Ai]，其中 Ai 是简单事件（即我们知道或能够轻易计算 P[Ai]）。交集 TiAi 对应于事件 Ai 的逻辑“与”，而并集 SiAi 对应于它们的逻辑“或”。例如，如果 Ai 表示某个系统中类型为 i 的故障发生，则 SiAi 就是系统发生故障的事件。

一般来说，计算此类组合的概率可能非常困难。在本节中，我们讨论一些可以进行计算的情形。让我们从独立事件开始，其交集的计算非常简单。

4.1 独立事件

定义 14.3 (独立性)。在同一概率空间中的两个事件 A,B 被称为是独立的，如果 P[A∩B] = P[A]×P[B]。

The intuition behind this definition is the following. Suppose that $\mathbb{P}[B] > 0$. Then we have

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} = \frac{\mathbb{P}[A] \times \mathbb{P}[B]}{\mathbb{P}[B]} = \mathbb{P}[A].$$

Thus independence has the natural meaning that "the probability of A is not affected by whether or not B occurs." (By a symmetrical argument, we also have $\mathbb{P}[B|A] = \mathbb{P}[B]$ provided $\mathbb{P}[A] > 0$.) For events A, B such that $\mathbb{P}[B] > 0$, the condition $\mathbb{P}[A|B] = \mathbb{P}[A]$ is actually *equivalent* to the definition of independence.

Several of our previously mentioned random experiments consist of independent events. For example, if we flip a coin twice, the event of obtaining heads in the first trial is independent of the event of obtaining heads in the second trial. The same applies for two rolls of a die; the outcomes of each trial are independent.

The above definition generalizes to any finite set of events:

Definition 14.4 (Mutual independence)**.** *Events $A_1, \ldots, A_n$ are said to be mutually independent if for **every** subset $I \subseteq \{1, \ldots, n\}$ with size $|I| \geq 2$,*

$$\mathbb{P}[\cap_{i \in I} A_i] = \prod_{i \in I} \mathbb{P}[A_i]. \tag{6}$$

An equivalent definition of mutual independence is as follows.

Definition 14.5 (Mutual independence)**.** *Events $A_1, \ldots, A_n$ are said to be mutually independent if for all $B_i \in \{A_i, \overline{A}_i\}, i = 1, \ldots, n$,*

$$\mathbb{P}[B_1 \cap \cdots \cap B_n] = \prod_{i=1}^{n} \mathbb{P}[B_i]. \tag{7}$$

Remarks.

1. In Definition 14.4, (6) needs to hold for *every* subset I of $\{1, \ldots, n\}$ with size $|I| \geq 2$. The cases of $|I| = 0$ and $|I| = 1$ are omitted, as they impose no constraints.

2. Note that (6) imposes $2^n - n - 1$ constraints on the probability distribution, while (7) defines 2^n constraints. It turns out that exactly $n + 1$ constraints implied by (7) are actually redundant.

For mutually independent events $A_1, \ldots, A_n$, it is not hard to check from the definition of conditional probability that, for any $1 \leq i \leq n$ and any subset $I \subseteq \{1, \ldots, n\} \setminus \{i\}$, we have

$$\mathbb{P}[A_i | \textstyle\bigcap_{j \in I} A_j] = \mathbb{P}[A_i].$$

Note that the independence of every pair of events (so-called *pairwise independence*) does *not* necessarily imply mutual independence. As illustrated in the following example, it is possible to construct three events A, B, C such that each *pair* is independent but the triple A, B, C is *not* mutually independent.

Example: Pairwise Independent but Not Mutually Independent

Suppose you toss a fair coin twice and let A be the event that the first flip is H and B be the event that the second flip is H. Now let C be the event that both flips are the same (i.e., both H's or both T's). Of course A and B are independent. What is more interesting is that so are A and C: given that the first toss came up H, the chance of the second flip being the same as the first is still $1/2$. Another way of saying this is that $\mathbb{P}[A \cap C] = \mathbb{P}[A]\mathbb{P}[C] = 1/4$ since $A \cap C$ is the event that the first flip is H and the second is also H. By the

此定义背后的直观含义如下。假设 P[B] > 0。那么我们有

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A\cap B]}{\mathbb{P}[B]} = \frac{\mathbb{P}[A]\times\mathbb{P}[B]}{\mathbb{P}[B]} = \mathbb{P}[A].$$

因此，独立性具有自然的意义：A 的概率不受 B 是否发生的影响。（通过对称论证，若 P[A] > 0，则也有 P[B|A] = P[B]。）对于满足 P[B] > 0 的事件 A,B，条件 P[A|B] = P[A] 实际上等价于独立性的定义。我们之前提到的多个随机实验都包含独立事件。例如，如果我们抛两次硬币，第一次试验中得到正面的事件与第二次试验中得到正面的事件是独立的。掷两次骰子的情况也是如此；每次试验的结果都是独立的。上述定义可推广到任意有限个事件：

定义 14.4 (相互独立性)。事件 A1,...,An 被称为是相互独立的，如果对于每一个子集 I ⊆{1,...,n} 且大小 |I| ⩾2,

$$\mathbb{P}[\cap_{i\in I}A_i] = \prod_{i\in I}\mathbb{P}[A_i]. \tag{6}$$

相互独立性的等价定义如下。

定义 14.5 (相互独立性)。事件 A1,...,An 被称为是相互独立的，如果对于所有 $B_i \in \{A_i, \overline{A}_i\}, i = 1,\ldots,n,$

$$\mathbb{P}[B_1\cap\cdots\cap B_n] = \prod_{i=1}^{n}\mathbb{P}[B_i]. \tag{7}$$

备注。

1. 在定义 14.4 中，(6) 必须对 {1,...,n} 的每一个大小 |I| ⩾2 的子集 I 成立。|I| = 0 和 |I| = 1 的情况被省略，因为它们不施加任何约束。
2. 注意 (6) 对概率分布施加了 2n −n−1 个约束，而 (7) 定义了 2n 个约束。事实上，(7) 所隐含的恰好有 n+1 个约束是冗余的。

对于相互独立的事件 A1,...,An，从条件概率的定义不难验证对于任意 1 ⩽i ⩽n 和任意子集 I ⊆{1,...,n}\{i}，都有 P[Ai|Tj∈I Aj]

= P[Ai] 的概率。

请注意，每对事件之间的独立性（所谓两两独立）并不一定意味着相互独立。如下例所示，可以构造三个事件 A、B、C，使得每一对都独立，但三者整体并不相互独立。

示例：两两独立但非相互独立

假设你抛一枚均匀硬币两次，令 A 表示第一次抛掷结果为正面的事件，B 表示第二次抛掷结果为正面的事件。现在令 C 表示两次抛掷结果相同的事件（即均为正面或均为反面）。显然 A 和 B 是独立的。更有趣的是，A 和 C 也是独立的：已知第一次抛掷为正面，第二次抛掷与第一次相同的概率仍然是 1/2。另一种说法是 P[A∩C] = P[A]P[C] = 1/4，因为 A∩C 是第一次为正面且第二次也为正面的事件。根据

same reasoning B and C are also independent. On the other hand, A, B and C are not mutually independent. For example, if we are given that A and B occurred, then the probability that C occurs is 1. So, even though A, B and C are not mutually independent, every pair of them are independent. In other words, A, B and C are pairwise independent but not mutually independent.

4.2 Intersections of Events

Computing the probability of an intersection of mutually independent events is easy; it follows from the definition. We simply multiply the probabilities of each event. How do we compute the probability of an intersection for events that are not mutually independent? From the definition of conditional probability, we immediately have the following Product Rule (sometimes also called the chain rule) for computing the probability of an intersection of events.

Theorem 14.1 (Product Rule)**.** *For any events A, B, we have*

$$\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B|A].$$

More generally, for any events $A_1, \ldots, A_n$,

$$\mathbb{P}[\textstyle\bigcap_{i=1}^{n} A_i] = \mathbb{P}[A_1] \times \mathbb{P}[A_2|A_1] \times \mathbb{P}[A_3|A_1 \cap A_2] \times \cdots \times \mathbb{P}[A_n|\bigcap_{i=1}^{n-1} A_i].$$

Proof. The first assertion follows directly from the definition of $\mathbb{P}[B|A]$ (and is in fact a special case of the second assertion with $n = 2$).

To prove the second assertion, we will use induction on n (the number of events). The base case is $n = 1$, and corresponds to the statement that $\mathbb{P}[A] = \mathbb{P}[A]$, which is trivially true. For an arbitrary $n > 1$, assume (the inductive hypothesis) that

$$\mathbb{P}[\textstyle\bigcap_{i=1}^{n-1} A_i] = \mathbb{P}[A_1] \times \mathbb{P}[A_2|A_1] \times \cdots \times \mathbb{P}[A_{n-1}|\bigcap_{i=1}^{n-2} A_i].$$

Now we can apply the definition of conditional probability to the two events A_n and $\bigcap_{i=1}^{n-1} A_i$ to deduce that

$$\begin{aligned}\mathbb{P}[\textstyle\bigcap_{i=1}^{n} A_i] = \mathbb{P}[A_n \cap (\bigcap_{i=1}^{n-1} A_i)] &= \mathbb{P}[A_n|\textstyle\bigcap_{i=1}^{n-1} A_i] \times \mathbb{P}[\bigcap_{i=1}^{n-1} A_i] \\ &= \mathbb{P}[A_n|\textstyle\bigcap_{i=1}^{n-1} A_i] \times \mathbb{P}[A_1] \times \mathbb{P}[A_2|A_1] \times \cdots \times \mathbb{P}[A_{n-1}|\bigcap_{i=1}^{n-2} A_i],\end{aligned}$$

where in the last line we have used the inductive hypothesis. This completes the proof by induction. □

The Product Rule is particularly useful when we can view our sample space as a sequence of choices. The next few examples illustrate this point.

Example: Coin Tosses

Toss a fair coin three times. Let A be the event that all three tosses are heads. Then $A = A_1 \cap A_2 \cap A_3$, where A_i is the event that the ith toss comes up heads. We have

$$\begin{aligned}\mathbb{P}[A] &= \mathbb{P}[A_1] \times \mathbb{P}[A_2|A_1] \times \mathbb{P}[A_3|A_1 \cap A_2] \\ &= \mathbb{P}[A_1] \times \mathbb{P}[A_2] \times \mathbb{P}[A_3] \\ &= \tfrac{1}{2} \times \tfrac{1}{2} \times \tfrac{1}{2} = \tfrac{1}{8}.\end{aligned}$$

The second line here follows from the fact that the tosses are mutually independent. Of course, we already know that $\mathbb{P}[A] = \frac{1}{8}$ from our definition of the probability space in the previous note. Another way of looking

同理，B 和 C 也是独立的。另一方面，A、B 和 C 并不相互独立。例如，如果已知 A 和 B 发生，则 C 发生的概率为 1。因此，尽管 A、B 和 C 不是相互独立的，但它们每一对都是独立的。换句话说，A、B 和 C 是两两独立的，但不是相互独立的。

4.2 事件的交集

计算相互独立事件交集的概率很容易；根据定义，只需将每个事件的概率相乘即可。那么如何计算非相互独立事件交集的概率呢？由条件概率的定义，我们立即得到以下用于计算事件交集概率的乘法规则（有时也称为链式法则）。

定理 14.1（乘法规则）。对于任意事件 A,B，有 P[A∩B] =

P[A]P[B|A]。更一般地，对于任意事件 A1,...,An，

$$\mathbb{P}[\bigcap_{i=1}^{n} A_i] = \mathbb{P}[A_1] \times \mathbb{P}[A_2|A_1] \times \mathbb{P}[A_3|A_1 \cap A_2] \times \cdots \times \mathbb{P}[A_n|\bigcap_{i=1}^{n-1} A_i]。$$

证明。第一个断言直接来自 P[B|A] 的定义（事实上是当 n = 2 时第二个断言的一个特例）。

为了证明第二个断言，我们将对 n（事件的数量）使用数学归纳法。基础情况是 n = 1，对应于 P[A] = P[A] 的陈述，这显然是成立的。对于任意的 n > 1，假设（归纳假设）

$$\mathbb{P}[\bigcap_{i=1}^{n-1} A_i] = \mathbb{P}[A_1] \times \mathbb{P}[A_2|A_1] \times \cdots \times \mathbb{P}[A_{n-1}|\bigcap_{i=1}^{n-2} A_i]。$$

现在我们可以将条件概率的定义应用于事件 An 和 Tn−1 i=1 Ai 来推导出

$$\begin{aligned}\mathbb{P}[\bigcap_{i=1}^{n} A_i] &= \mathbb{P}[A_n \cap (\bigcap_{i=1}^{n-1} A_i)] = \mathbb{P}[A_n|\bigcap_{i=1}^{n-1} A_i] \times \mathbb{P}[\bigcap_{i=1}^{n-1} A_i] \\ &= \mathbb{P}[A_n|\bigcap_{i=1}^{n-1} A_i] \times \mathbb{P}[A_1] \times \mathbb{P}[A_2|A_1] \times \cdots \times \mathbb{P}[A_{n-1}|\bigcap_{i=1}^{n-2} A_i],\end{aligned}$$

其中最后一行使用了归纳假设。这就完成了归纳证明。 □

当我们可以将样本空间视为一系列选择时，乘法规则特别有用。接下来的几个例子说明了这一点。

示例：抛硬币

抛一枚均匀硬币三次。设 A 为三次都出现正面的事件。则 A = A1 ∩A2 ∩A3，其中 Ai 是第 i 次抛掷结果为正面的事件。我们有

$$\begin{aligned}\mathbb{P}[A] &= \mathbb{P}[A_1] \times \mathbb{P}[A_2|A_1] \times \mathbb{P}[A_3|A_1 \cap A_2] \\ &= \mathbb{P}[A_1] \times \mathbb{P}[A_2] \times \mathbb{P}[A_3] \\ &= \tfrac{1}{2} \times \tfrac{1}{2} \times \tfrac{1}{2} = \tfrac{1}{8}.\end{aligned}$$

这里的第二行源于抛掷相互独立的事实。当然，我们已经知道 P[A] = 1
8 从上一讲义中概率空间的定义来看，另一种观察方式

at this calculation is that it justifies our definition of the probability space, and shows that it was consistent with assuming that the coin flips are mutually independent.

If the coin is biased with heads probability p, we get, again using independence,

$$\mathbb{P}[A] = \mathbb{P}[A_1] \times \mathbb{P}[A_2] \times \mathbb{P}[A_3] = p^3.$$

More generally, the probability of any sequence of n tosses containing k heads and $n-k$ tails is $p^k(1-p)^{n-k}$. This is in fact the reason we defined the probability space this way: we defined the sample point probabilities so that the coin tosses would behave independently.

Example: Monty Hall Revisited

Recall the Monty Hall problem from the previous note: there are three doors and the probability that the prize is behind any given door is $\frac{1}{3}$. There are goats behind the other two doors. The contestant picks a door randomly, and the host opens one of the other two doors, revealing a goat. How do we calculate intersections in this setting? For example, what is the probability that the contestant chooses door 1, the prize is behind door 2, and the host chooses door 3?

Let C_i be the event that the contestant chooses door i, let P_i be the event that the prize is behind door i, and let H_i be the event that the host chooses door i. Then, by the Product Rule,

$$\mathbb{P}[C_1 \cap P_2 \cap H_3] = \mathbb{P}[C_1] \times \mathbb{P}[P_2|C_1] \times \mathbb{P}[H_3|C_1 \cap P_2].$$

The probability of C_1 is $\frac{1}{3}$, since the contestant is choosing the door at random. The probability of P_2 given C_1 is still $\frac{1}{3}$ since they are independent. The probability of the host choosing door 3 given events C_1 and P_2 is 1; the host cannot choose door 1 since the contestant has already chosen it, and the host cannot choose door 2 since the host must reveal a goat (and not the prize). Therefore,

$$\mathbb{P}[C_1 \cap P_2 \cap H_3] = \frac{1}{3} \times \frac{1}{3} \times 1 = \frac{1}{9}.$$

Observe that we needed conditional probability in this setting; had we simply multiplied the probabilities of each event, we would have obtained $\frac{1}{27}$ since the probability of H_3 is also $\frac{1}{3}$ (can you figure out why?). What if we changed the situation, and instead asked for the probability that the contestant chooses door 1, the prize is behind door 1, and the host chooses door 3? We can use the same technique as above, but our final answer will be different. This is left as an exercise (see Exercise 4).

Now, noting that $\mathbb{P}[H_3|C_1] = \frac{1}{2}$ (see Exercise 5) and $\mathbb{P}[C_1 \cap H_3] = \mathbb{P}[C_1]\,\mathbb{P}[H_3|C_1] = \frac{1}{3} \times \frac{1}{2} = \frac{1}{6}$, we obtain

$$\mathbb{P}[P_2 \mid C_1 \cap H_3] = \frac{\mathbb{P}[C_1 \cap P_2 \cap H_3]}{\mathbb{P}[C_1 \cap H_3]} = \frac{\frac{1}{9}}{\frac{1}{6}} = \frac{2}{3},$$

which formally justifies the intuitive answer described in the previous note.

Example: Poker Hands

Let's use the Product Rule to compute the probability of a flush in a different way. This is equal to $4 \times \mathbb{P}[A]$, where A is the probability of a Hearts flush. Intuitively, this should be clear since there are 4 suits; we'll see

这种计算的意义在于，它验证了我们对概率空间的定义，并表明该定义与假设硬币翻转相互独立是一致的。

如果硬币是有偏的，正面概率为 p，再次利用独立性，可得 P[A] =

P[A1]×P[A2]×P[A3] = p3。

更一般地，任何包含 k 次正面和 n−k 次反面的 n 次投掷序列的概率为 pk(1−p)n−k。这实际上是我们如此定义概率空间的原因：我们定义样本点的概率，使得每次抛硬币相互独立。

示例：再谈蒙提霍尔问题

回顾上一讲义中的蒙提霍尔问题：共有三扇门，奖品在任意一扇门后的概率都是 1
3 另外两扇门后面是山羊。参赛者选择一扇门
随机地，主持人打开另外两扇门中的一扇，露出一只山羊。在这种情况下我们如何计算交集？例如，参赛者选择第 1 扇门、奖品在第 2 扇门后且主持人选择第 3 扇门的概率是多少?

设 Ci 表示参赛者选择第 i 扇门的事件，Pi 表示奖品在第 i 扇门后的事件，Hi 表示主持人选择第 i 扇门的事件。那么根据乘法规则，

$$\mathbb{P}[C_1 \cap P_2 \cap H_3] = \mathbb{P}[C_1] \times \mathbb{P}[P_2|C_1] \times \mathbb{P}[H_3|C_1 \cap P_2].$$

C1的概率是13，因为参赛者是随机选择门的。在C1发生的条件下P2的概率仍然是1
3 因为它们相互独立。在事件C1和P2发生的条件下，主持人选择3号门的概率是1；主持人不能选择1号门，因为参赛者已经选择了，也不能选择2号门，因为主持人必须揭示一只山羊（而不是奖品）。因此，

$$\mathbb{P}[C_1 \cap P_2 \cap H_3] = \frac{1}{3} \times \frac{1}{3} \times 1 = \frac{1}{9}.$$

请注意，这种情况我们需要条件概率；如果只是简单地将每个事件的概率相乘，我们会得到1
27 因为H3的概率也是1 3 （你能想出原因吗？）
如果我们改变情境，转而询问参赛者选择1号门、奖品在1号门后、主持人选择3号门的概率是多少？我们可以使用与上面相同的技术，但最终答案会不同。这留作练习（见练习4）。

现在注意到P[H3|C1] = 1（见练习5），且P[C1 ∩H3] = P[C1]P[H3|C1] = 1 ×1 3 = 1 2 6，我们得到

$$\mathbb{P}[P_2 \mid C_1 \cap H_3] = \frac{\mathbb{P}[C_1 \cap P_2 \cap H_3]}{\mathbb{P}[C_1 \cap H_3]} = \frac{\frac{1}{9}}{\frac{1}{6}} = \frac{2}{3},$$

这正式验证了前一篇笔记中描述的直观答案。

示例：扑克牌型

让我们用乘法法则以另一种方式计算同花的概率。这等于4×P[A]，其中A表示红桃花色同花的概率。直观上这应该是清楚的，因为共有4种花色；我们将在后面看到

why this is formally true in the next section. We can write $A = \bigcap_{i=1}^{5} A_i$, where A_i is the event that the ith card we pick is a Heart. So we have

$$\mathbb{P}[A] = \mathbb{P}[A_1] \times \mathbb{P}[A_2|A_1] \times \cdots \times \mathbb{P}[A_5|\bigcap_{i=1}^{4} A_i].$$

Clearly $\mathbb{P}[A_1] = \frac{13}{52} = \frac{1}{4}$. What about $\mathbb{P}[A_2|A_1]$? Well, since we are conditioning on A_1 (the first card is a Heart), there are only 51 remaining possibilities for the second card, 12 of which are Hearts. So $\mathbb{P}[A_2|A_1] = \frac{12}{51}$. Similarly, $\mathbb{P}[A_3|A_1 \cap A_2] = \frac{11}{50}$, and so on. So we get

$$4 \times \mathbb{P}[A] = 4 \times \frac{13}{52} \times \frac{12}{51} \times \frac{11}{50} \times \frac{10}{49} \times \frac{9}{48},$$

which is exactly the same fraction we computed in the previous note.

So now we have two methods of computing probabilities in many of our sample spaces. It is useful to keep these different methods around, both as a check on your answers and because in some cases one of the methods is easier to use than the other.

4.3 Unions of Events

You are in Las Vegas, and you spy a new game with the following rules. You pick a number between 1 and 6. Then three dice are thrown. You win if and only if your number comes up on at least one of the dice.

The casino claims that your odds of winning are 50%, using the following argument. Let A be the event that you win. We can write $A = A_1 \cup A_2 \cup A_3$, where A_i is the event that your number comes up on die i. Clearly $\mathbb{P}[A_i] = \frac{1}{6}$ for each i. Therefore, they claim

$$\mathbb{P}[A] = \mathbb{P}[A_1 \cup A_2 \cup A_3] = \mathbb{P}[A_1] + \mathbb{P}[A_2] + \mathbb{P}[A_3] = 3 \times \frac{1}{6} = \frac{1}{2}.$$

Is this calculation correct? Well, suppose instead that the casino rolled six dice, and again you win if and only if your number comes up at least once. Then the analogous calculation would say that you win with probability $6 \times \frac{1}{6} = 1$, i.e., certainly! The situation becomes even more ridiculous when the number of dice gets bigger than 6.

The problem is that the events A_i are *not disjoint*: i.e., there are some sample points that lie in more than one of the A_i. (We could get really lucky and our number could come up on two of the dice, or all three.) So, if we add up the $\mathbb{P}[A_i]$, we are counting some sample points more than once.

Fortunately, in Note 11 we learned about the Principle of Inclusion-Exclusion which allows us to deal with this kind of situation. In the proof of Theorem 11.3, it was shown that every $\omega \notin A_1 \cup \cdots \cup A_n$ is not counted by the Inclusion-Exclusion formula, while every $\omega \in A_1 \cup \cdots \cup A_n$ is counted exactly once by the formula. Hence, rather than summing (with appropriate signs) the cardinality of $\cap_{i \in S} A_i$ in the Inclusion-Exclusion formula, if we instead sum their probability $\mathbb{P}[\cap_{i \in S} A_i]$, we obtain the following useful formula for computing $\mathbb{P}[A_1 \cup \cdots \cup A_n]$:

Theorem 14.2 (Inclusion-Exclusion)**.** *Let $A_1, \ldots, A_n$ be events in some probability space, where $n \geq 2$. Then, we have*

$$\mathbb{P}[A_1 \cup \cdots \cup A_n] = \sum_{k=1}^{n} (-1)^{k-1} \sum_{S \subseteq \{1,\ldots,n\}:|S|=k} \mathbb{P}[\cap_{i \in S} A_i]. \tag{8}$$

The right hand side of (8) can be written as

$$\mathbb{P}[\bigcup_{i=1}^{n} A_i] = \sum_{i=1}^{n} \mathbb{P}[A_i] - \sum_{i<j} \mathbb{P}[A_i \cap A_j] + \sum_{i<j<k} \mathbb{P}[A_i \cap A_j \cap A_k] - \cdots + (-1)^{n-1} \mathbb{P}[A_1 \cap A_2 \cap \cdots \cap A_n],$$

为什么这在形式上成立将在下一节说明。我们可以将A写成T5i=1Ai，其中Ai表示我们抽到的第i张牌是红桃的事件。因此我们有

$$\mathbb{P}[A] = \mathbb{P}[A_1] \times \mathbb{P}[A_2|A_1] \times \cdots \times \mathbb{P}[A_5|\bigcap_{i=1}^{4} A_i].$$

显然P[A1] = 13 $\frac{}{52}$ = 14。那P[A2|A1]呢？由于我们是在A1（第一张牌是红桃），第二张牌只剩下51种可能，其中有12张是红桃。因此P[A2|A1] = 12 51。类似地，P[A3|A1 ∩A2] = 11，依此类推。

所以我们得到 4×P[A] = 4× 13 52 $\quad — \times \frac{12}{51} \times \frac{11}{50} \times \frac{10}{49} \times \frac{9}{48},$

这正是我们在前一篇笔记中计算出的相同比例。

现在我们在许多样本空间中有两种计算概率的方法。保留这些不同方法是有用的，既可以用来检验答案，也因为在某些情况下其中一种方法比另一种更容易使用。

4.3 事件的并集

你在拉斯维加斯，发现一种新游戏规则如下：你从1到6之间选一个数字，然后掷三个骰子。只要你选的数字至少在一个骰子上出现，你就赢了。赌场声称你的胜率是50%，其理由如下。令A表示你赢的事件。我们可以写成A = A1 ∪A2 ∪A3，其中Ai表示你的数字出现在第i个骰子上的事件。显然P[Ai] = 1 $\frac{}{6}$ 对每个i成立。因此他们声称

$$\mathbb{P}[A] = \mathbb{P}[A_1 \cup A_2 \cup A_3] = \mathbb{P}[A_1] + \mathbb{P}[A_2] + \mathbb{P}[A_3] = 3 \times \frac{1}{6} = \frac{1}{2}.$$

这个计算正确吗？假设赌场改为掷六个骰子，你仍然只要号码至少出现一次就算赢。那么类似的计算会得出你获胜的概率为6×1 $\frac{}{6}$ = 1，即必然发生！当骰子数量增加时，情况变得更加荒谬。超过6。

问题在于事件Ai并不互斥：即存在一些样本点落在多个Ai中。（我们可能非常幸运，我们的号码同时出现在两个骰子上，甚至三个都出现。）因此，如果我们把所有的P[Ai]加起来，就会重复计算某些样本点。

幸运的是，在笔记11中我们学习了容斥原理，它使我们能够处理这类情况。在定理11.3的证明中，已经表明每一个ω/∈A1 ∪···∪An都不会被计入根据容斥公式，而每一个ω∈A1 ∪···∪An在公式中恰好被计算一次。因此，在容斥公式中，与其对∩i∈SAi的基数进行带符号求和，不如改为对其概率P[∩i∈SAi]求和，从而得到一个用于计算P[A1 ∪···∪An]的有用公式：

定理14.2（容斥原理）。设A1,...,An是某个概率空间中的事件，其中n ⩾2。那么，我们有

$$\mathbb{P}[A_1 \cup \cdots \cup A_n] = \sum_{k=1}^{n} (-1)^{k-1} \sum_{S \subseteq \{1,\ldots,n\}:|S|=k} \mathbb{P}[\cap_{i \in S} A_i]. \tag{8}$$

式(8)的右边可以写成

$$\mathbb{P}[\bigcup_{i=1}^{n} A_i] = \sum_{i=1}^{n} \mathbb{P}[A_i] - \sum_{i<j} \mathbb{P}[A_i \cap A_j] + \sum_{i<j<k} \mathbb{P}[A_i \cap A_j \cap A_k] - \cdots + (-1)^{n-1} \mathbb{P}[A_1 \cap A_2 \cap \cdots \cap A_n],$$

where $\sum_{i<j}$ denotes summing over all $i, j \in \{1, \ldots, n\}$ such that $i < j$, and so on. That is, to compute $\mathbb{P}[\bigcup_i A_i]$, we start by summing the event probabilities $\mathbb{P}[A_i]$, then we *subtract* the probabilities of all pairwise intersections, then we *add* back in the probabilities of all three-way intersections, and so on. You might like to verify it for the special case $n = 3$ by drawing a Venn diagram. You might also like to prove the formula for general n by induction (in similar fashion to the proof of the Product Rule above).

Using (8), the probability we get lucky in the new game in Las Vegas is given by

$$\mathbb{P}[A_1 \cup A_2 \cup A_3] = \mathbb{P}[A_1] + \mathbb{P}[A_2] + \mathbb{P}[A_3] - \mathbb{P}[A_1 \cap A_2] - \mathbb{P}[A_1 \cap A_3] - \mathbb{P}[A_2 \cap A_3] + \mathbb{P}[A_1 \cap A_2 \cap A_3].$$

Now the nice thing here is that the events A_i are mutually independent (the outcome of any die does not depend on that of the others), so $\mathbb{P}[A_i \cap A_j] = \mathbb{P}[A_i]\mathbb{P}[A_j] = (\frac{1}{6})^2 = \frac{1}{36}$, and similarly $\mathbb{P}[A_1 \cap A_2 \cap A_3] = (\frac{1}{6})^3 = \frac{1}{216}$. So, we get

$$\mathbb{P}[A_1 \cup A_2 \cup A_3] = \left(3 \times \tfrac{1}{6}\right) - \left(3 \times \tfrac{1}{36}\right) + \tfrac{1}{216} = \tfrac{91}{216} \approx 0.42.$$

So your odds are quite a bit worse than what the casino is claiming!

When n is large (i.e., we are interested in the union of many events), the Inclusion-Exclusion formula is essentially useless because it involves computing the probability of the intersection of every non-empty subset of the events: and there are $2^n - 1$ of these! Sometimes we can just look at the first few terms of it and forget the rest: note that successive terms actually give us an overestimate and then an underestimate of the answer, and these estimates both get better as we go along.

However, in many situations we can get a long way by just looking at the first term:

1. (**Mutually exclusive events**) If the events $A_1, \ldots, A_n$ are mutually exclusive (i.e., $A_i \cap A_j = \varnothing$ for all $i \neq j$), then

   $$\mathbb{P}[\textstyle\bigcup_{i=1}^n A_i] = \displaystyle\sum_{i=1}^n \mathbb{P}[A_i].$$

   [Note that we have already used this fact several times in our examples, e.g., in claiming that the probability of a flush is four times the probability of a Hearts flush — clearly flushes in different suits are disjoint events.]

2. (**Union bound**) Let $A_1, \ldots, A_n$ be events in some probability space. Then, for all $n \in \mathbb{Z}^+$,

   $$\mathbb{P}[\textstyle\bigcup_{i=1}^n A_i] \leq \displaystyle\sum_{i=1}^n \mathbb{P}[A_i]. \tag{9}$$

   This merely says that adding up the $\mathbb{P}[A_i]$ can only *over*estimate the probability of the union. Crude as it may seem, we will later see how to use the union bound effectively in Computer Science examples.

5 Exercises

1. Suppose five cards are dealt from a standard deck of playing cards. Let A_i be the event that the ith card is an ace. Show that $\mathbb{P}[A_i] = \frac{1}{13}$ for all $i = 1, \ldots, 5$.
2. Show (4) and (5).
3. Show that Definitions 14.4 and 14.5 of mutual independence are equivalent.
4. In the Monty Hall problem, find $\mathbb{P}[C_1 \cap P_1 \cap H_3]$.

其中 Σi<j 表示对所有满足 i < j 的 i, j ∈{1,...,n} 求和，依此类推。也就是说，为了计算 P[SiAi]，我们先对事件概率 P[Ai] 求和，然后减去所有两两交集的概率，再加回所有三重交集的概率，如此反复。你可以通过画韦恩图来验证 n = 3 的特殊情况。你也可以尝试用归纳法来证明一般情况下的公式（方法类似于上面乘积法则的证明）。

使用（8），我们在拉斯维加斯的新游戏中获胜的概率为

P[A1 ∪A2 ∪A3] = P[A1]+P[A2]+P[A3]−P[A1 ∩A2]−P[A1 ∩A3]−P[A2 ∩A3]+P[A1 ∩A2 ∩A3]。

现在这里的好消息是，事件 Ai 是相互独立的（任意骰子的结果不依赖于其他骰子），因此 P[Ai ∩Aj] = P[Ai]P[Aj] = (1 $\frac{}{6})^2 = \frac{1}{36}$, 以及类似地 P[A1 ∩A2 ∩A3] = $(\frac{1}{6})$3 = 1216。因此，我们得到

$$\mathbb{P}[A_1 \cup A_2 \cup A_3] = \left(3 \times \tfrac{1}{6}\right) - \left(3 \times \tfrac{1}{36}\right) + \tfrac{1}{216} = \tfrac{91}{216} \approx 0.42。$$

所以你的胜算比赌场宣称的要差得多！

当 n 很大时（即我们关注的是大量事件的并集），容斥公式实际上变得毫无用处，因为它需要计算每个非空子集事件交集的概率：而这样的子集有 2n −1 个！有时我们只需看前几项，然后忽略其余部分：请注意，连续的项实际上交替给出过高估计和过低估计，且这些估计随着项数增加而变得更好。

然而，在许多情况下，我们只需看第一项就可取得很大进展：

1. （互斥事件）如果事件 A1,...,An 是互斥的（即对所有 i = j，有 Ai ∩Aj = ∅），则

$$\mathbb{P}[\textstyle\bigcup_{i=1}^n A_i] = \sum_{i=1}^{n} \text{P[Ai]。}$$

［注意，我们已经在多个例子中使用过这一事实，例如，声称同花的概率是红心同花的四倍——显然不同花色的同花是互斥事件。］

2. （并界）设 A1,...,An 是某个概率空间中的事件。那么，对于所有 n ∈Z+，

$$\mathbb{P}[\textstyle\bigcup_{i=1}^n A_i] \le \sum_{i=1}^{n} \text{P[Ai]。} \tag{9}$$

这仅仅说明，将 P[Ai] 相加只会高估并集的概率。尽管看起来粗糙，我们之后会看到如何在计算机科学的例子中有效使用并界。

5 练习

1. 假设从标准扑克牌组中发出五张牌。令 Ai 表示第 i 张牌是 A 的事件。证明 P[Ai] = 1 $\frac{}{13}$ 对所有 i = 1,...,5。
2. 证明（4）和（5）。
3. 证明互独立性的定义14.4和14.5是等价的。
4. 在蒙提霍尔问题中，求 P[C1 ∩P1 ∩H3]。

5. In the Monty Hall problem, show that $\mathbb{P}[H_3 \mid C_1] = \frac{1}{2}$.
6. Prove the union bound (9).

5. 在蒙提霍尔问题中，证明 $P[H3 \mid C1] = \frac{1}{2}$.

6. 证明并界（9）。