

Contents

1	条件概率的一般形式	2
1.1	简要复习	2
1.2	例1: 古典情形中的条件概率	4
1.3	例2: 一般情形中的条件概率	5
1.4	贝叶斯定理	6
2	决策树	7
2.1	一个引入例子: 攀岩者	7
2.2	例子: 丛林和小偷	8
2.3	回到攀岩者	10
2.4	练习: 不放回抽取	11
3	著名例子	13
3.1	蒙提霍尔问题	13
3.2	生日悖论	14
3.3	医学检测	16

在本讲中, 我们将发展 条件概率。正如我们在第5讲中简要看到的, 条件概率是指在我们获得了关于所研究系统的某种额外信息之后所计算的概率。现在, 我们将通过考察古典概率之外的一些一般计算, 更仔细地研究这一概念。我们还会介绍 决策树, 它为组织多步骤概率问题提供了一种实用方法。最后, 我们将讨论三个涉及条件概率的著名例子: 蒙提霍尔问题、生日悖论和医学检测。

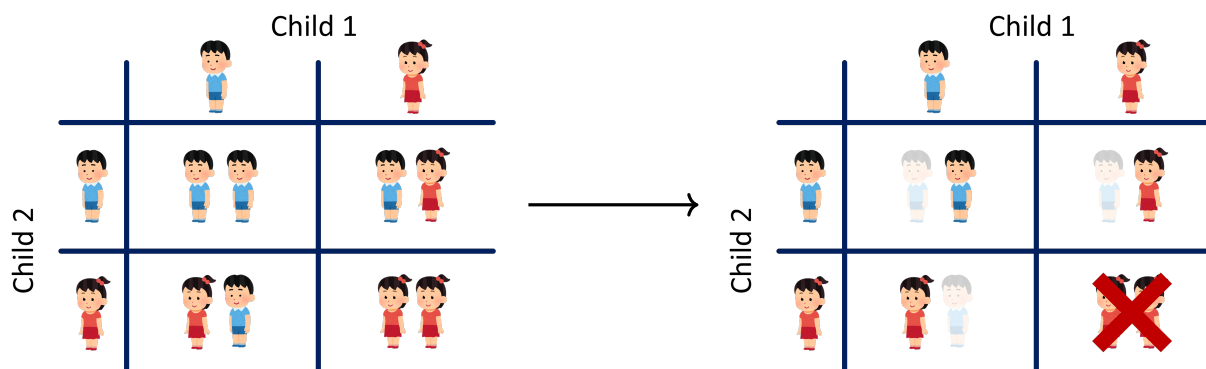
1 条件概率的一般形式

假设你知道你的邻居有两个孩子。你去他们家敲门，一个男孩来开门。那么，另一个孩子也是男孩的概率是多少？

令人惊讶的是，答案其实是 $\frac{1}{3}$ 。第一次听到这个答案时，它确实相当令人意外，但它是正确的。我们简单分析一下这个问题。首先，可能的有序结果有四个：

$$S = \{BB, BG, GB, GG\}.$$

朴素地说，我们会认为这些结果各自有 $\frac{1}{4}$ 的概率发生。然而，通过看到至少有一个孩子是男孩，我们获得了一次信息更新：现在我们知道“GG”这个选项已经不可能了。这把可能性的空间缩小为：



换句话说，样本空间实际上被缩小为：

$$\{(\text{男孩}, \text{男孩}), (\text{女孩}, \text{男孩}), (\text{男孩}, \text{女孩})\}.$$

这三个结果中，只有一个结果是两个孩子都是男孩。因此：

$$P(\text{两个男孩} \mid \text{至少一个男孩}) = \frac{1}{3}.$$

这就是一次条件概率计算。条件概率根据更新后的信息或额外假设来计算不确定性。正如你所看到的，条件概率的结果常常令人惊讶，并且违背直觉。今天我们将介绍几种方法，用来处理这种容易让人混乱的情况。我们会先简要复习前几讲中已经见过的一些理论，然后把它应用到一些基本计算中。之后，我们会学习决策树，它可以把概率整理成一个非常容易理解的结果系统。最后，我们将以几个涉及条件概率的著名反直觉例子结束本讲：蒙提霍尔问题、生日悖论，以及医学检测的可靠性。

1.1 简要复习

回忆一下，在古典概率中，所有单个结果都是等可能的。正如我们在第6讲中讨论过的，这是一种特殊假设，它给了我们一种确定概率的特殊方法：

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

上面的公式说明，由于每个结果都被假设为等可能，我们可以通过数出任意事件 A 的大小，并计算它在样本空间 S 中所占的比例，来确定 $P(A)$ 。

相反，在一般概率中，我们放松了“每个结果都有相同概率发生”这一假设。用 ω 表示一个单独的结果，在一般概率中，我们不假设所有概率 $p(\omega)$ 都相同。因此，我们不能仅仅通过数事件 A 的大小来确定它的概率。相反，我们必须把 A 中所有结果的概率加起来。这个和就是概率 $P(A)$ ：

$$P(A) = \sum_{\omega \text{ in } A} p(\omega).$$

正如我们在上一讲中提到的，复合事件的公式在一般情形下仍然有效。不过，它们的理由有所不同，但仍然可以照常使用。

1. 要确定事件 A 不发生的概率，可以使用公式：

$$P(\text{not } A) = 1 - P(A).$$

2. 要确定两个事件 A 和 B 同时发生的概率，可以使用以下公式：

$$P(A \text{ and } B) = \begin{cases} P(A) \cdot P(B), & \text{如果 } A \text{ 和 } B \text{ 独立,} \\ P(A | B) \cdot P(B), & \text{如果 } A \text{ 和 } B \text{ 相关, 并且 } P(B) \neq 0. \end{cases}$$

3. 要确定两个事件 A 或 B 至少一个发生，或两者都发生的概率，可以使用以下公式：

$$P(A \text{ or } B) = \begin{cases} P(A) + P(B), & \text{如果 } A \text{ 和 } B \text{ 互斥,} \\ P(A) + P(B) - P(A \text{ and } B), & \text{如果 } A \text{ 和 } B \text{ 不互斥.} \end{cases}$$

我们在第5讲中提到，符号 $P(A | B)$ 表示在我们知道 B 一定会发生的条件下， A 发生的概率。在那次讨论中，我们提到，对于古典概率，可以用 $P(A \text{ and } B)$ 的公式来发现 $P(A | B)$ 的公式。回忆一下，如果我们刻意把 $1 = \frac{|B|}{|B|}$ 插入到古典版的 Golden Formula 中，就会得到：

$$P(A \text{ and } B) = \frac{|A \cap B|}{|S|} = \frac{|A \cap B|}{|S|} \cdot \frac{|B|}{|B|} = \frac{|A \cap B|}{|B|} \cdot \frac{|B|}{|S|}.$$

第二个因子正是 $P(B)$ ，而第一个因子给出了一个新的公式：

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

用概率语言来说，这个公式是在把 B 当作样本空间，而不是从 S 开始时，确定 A 的概率。换句话说，它是在 B 被保证发生的条件下， A 的概率。

在一般概率论中，也有类似公式：

$$P(A | B) = \frac{\sum_{\omega \in A \cap B} p(\omega)}{\sum_{\omega \in B} p(\omega)} = \sum_{\omega \in A \cap B} \frac{p(\omega)}{P(B)}, \quad \text{provided } P(B) \neq 0.$$

这个公式看起来相当不漂亮。现在，只要你认识到它存在就足够了。过一会儿，我们会用一个清楚例子把它解释明白。

无论是古典情形还是一般情形，我们总是可以重新整理 $P(A \text{ and } B)$ 的公式，从而得到条件概率的另一种表达：

$$P(A \text{ and } B) = P(A | B) \cdot P(B) \longrightarrow P(A | B) = \frac{P(A \text{ and } B)}{P(B)}, \quad P(B) \neq 0.$$

如果我们交换 A 和 B 的角色，同样的过程也适用：

$$P(B \text{ and } A) = P(B | A) \cdot P(A) \longrightarrow P(B | A) = \frac{P(A \text{ and } B)}{P(A)}, \quad P(A) \neq 0.$$

这些公式给了我们一种求条件概率的有用方法：如果我们知道 $P(A \text{ and } B)$ ，以及我们假设已经知道的那个事件的概率，那么就可以使用上面的两个公式之一来计算条件概率。下面我们通过两个例子来说明这一点。

1.2 例1：古典情形中的条件概率

在第5讲中，我们已经简要看过条件概率。当时，我们接触了下面这个例子。假设你掷一个公平的 dice，并考虑两个事件：

$$A = \{\text{dice 掷出偶数}\},$$

以及

$$B = \{\text{dice 掷出小于 4 的数}\}.$$

由于我们使用的是普通的 6 面 dice，样本空间以及事件 A 和 B 可以用集合记法写成：

$$S = \{1, 2, 3, 4, 5, 6\}, \quad A = \{2, 4, 6\}, \quad B = \{1, 2, 3\}.$$

现在，想象你把 dice 掷到空中，它正在旋转时，时间突然慢了下来，神从天上降临。他在你耳边低声说：“dice 会落在一个小于 4 的数上。”在这一刻，你绝对知道事件 B 会发生。因此，你对事件 A 是否发生的不确定性也可能发生变化。

为了求出在 B 被保证发生时 A 发生的新概率，我们可以把整个样本空间有效地缩小到小于 4 的那部分数：

$$S = \{1, 2, 3, \cancel{4}, \cancel{5}, \cancel{6}\} \longrightarrow \{1, 2, 3\}$$

在这个受限制的系统中，我们注意到现在只有三个可能结果，而其中只有一个是偶数。因此，掷出偶数的概率会从原来的 $\frac{1}{2}$ 变为 $\frac{1}{3}$ 。我们刚刚做的事情，就是计算事件 A 在给定 B 条件下的条件概率。用符号表示，条件概率 $P(A | B)$ 可以通过重新整理 A 和 B 同时发生的概率公式得到：

$$P(A \text{ and } B) = P(A | B) \cdot P(B) \longrightarrow P(A | B) = \frac{P(A \text{ and } B)}{P(B)}.$$

通过观察可以看到，集合 A 与 B 的交集是 $A \cap B = \{2\}$ 。因此，上面的论证可以用符号写成：

$$P(A | B) = \frac{|A \cap B|}{|B|} = \frac{|\{2\}|}{|\{1, 2, 3\}|} = \frac{1}{3}.$$

或许更简单的方法是先直接计算 $P(A \text{ and } B)$ 和 $P(B)$ ，然后代入公式：

$$P(A | B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{1}{6} \div \frac{3}{6} = \frac{1}{3}.$$

1.3 例2：一般情形中的条件概率

现在假设我们把这个 dice 换成一个不公平的 dice，新的概率如下：

$$p(1) = 0.1, \quad p(2) = 0.2, \quad p(3) = 0.1, \quad p(4) = 0.1, \quad p(5) = 0.1, \quad p(6) = 0.4.$$

仍然使用前面的事件和样本空间：

$$S = \{1, 2, 3, 4, 5, 6\}, \quad A = \{2, 4, 6\}, \quad B = \{1, 2, 3\}.$$

现在注意，我们不是在使用古典概率。这意味着，我们不能简单地数子集的大小来确定概率。相反，我们必须把每个事件中所包含的各个结果的概率加起来。因此， A 和 B 的概率现在是：

$$B = \{1, 2, 3\} \implies P(B) = p(1) + p(2) + p(3) = 0.1 + 0.2 + 0.1 = 0.4$$

$$A = \{2, 4, 6\} \implies P(A) = p(2) + p(4) + p(6) = 0.2 + 0.1 + 0.4 = 0.7$$

因此，在这个不公平系统中，每次掷 dice 时，得到偶数的概率是 70%。现在，再次假设当我们把 dice 掷到空中时，神降临并告诉我们结果一定会是一个小于 4 的数。在这种情况下，我们对事件 A 的不确定性再次发生变化，因为我们已经排除了得到 4、5 或 6 的可能性。因此，基于给定的信息，新的系统有一个缩小后的样本空间：

$$S = \{1, 2, 3, \cancel{4}, \cancel{5}, \cancel{6}\} \longrightarrow \{1, 2, 3\}$$

粗略地说，结果 1, 2, 3 的概率现在需要被调整，使它们在这个新系统中加起来等于 1。数学上，做到这一点的方法是把每个概率除以缩小后样本空间的概率 $P(B) = 0.4$ 。注意，这样做之后，所有可能结果的和确实等于 1：

$$\frac{p(1)}{P(B)} + \frac{p(2)}{P(B)} + \frac{p(3)}{P(B)} = \frac{0.1}{0.4} + \frac{0.2}{0.4} + \frac{0.1}{0.4} = \frac{0.4}{0.4} = 1.$$

因此， B 中各个单独结果的条件概率，可以通过把原始概率除以缩小后样本空间的总概率来得到。特别地，如果我们看得到 2 的新概率，就有：

$$p(2 | B) = \frac{p(2)}{P(B)} = \frac{0.2}{0.4} = \frac{1}{2}.$$

这里， $p(2 | B)$ 是 $P(\{2\} | B)$ 的简写。从某种意义上说，这很合理。注意，在原来的系统中，三个结果 1, 2, 3 的概率分别是 0.1, 0.2, 0.1。得到 1 或 3 的概率相等，都是 0.1，而得到 2 的概率是它们的两倍，即 0.2。所以，如果我们突然只关注这三个结果，就会期望这种比例关系仍然保持： $P(2 | B)$ 应该是 $P(1 | B)$ 和 $P(3 | B)$ 的两倍，并且后两者应该相等。实现这一点的唯一方法就是：

$$P(1 | B) = \frac{1}{4}, \quad P(2 | B) = \frac{1}{2}, \quad \text{and} \quad P(3 | B) = \frac{1}{4}.$$

为了在这个系统中计算条件概率 $P(A | B)$ ，我们可以先计算 $P(A \text{ and } B)$ ，然后使用和前面相同的公式。要强调的是，我们不是在古典概率中，所以为了确定复合事件 $A \text{ and } B$ 的概率，我们需要查看交集 $A \cap B$ ，并把这个集合中每个结果的概率加起来。由于 $A \cap B = \{2\}$ ，这里只有一个结果，也就是掷出 2，因此这个和中只有一项。于是，我们得到 $P(A \text{ and } B) = 0.2$ 。

现在，我们可以使用条件概率公式来确认前面的论证：

$$P(A | B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{0.2}{0.4} = 0.5,$$

这与前面的答案相同。

练习

计算这个不公平 dice 的条件概率 $P(B | A)$ 。

解答

我们要计算 $P(B | A)$ ，也就是在已经知道 dice 落在偶数上的条件下，它落在小于 4 的数上的概率。交集是

$$A \cap B = \{2\},$$

所以和前面一样， $P(A \text{ and } B) = 0.2$ 。不过现在，在应用公式之前，我们需要先计算 $P(A)$ 。我们有：

$$P(A) = p(2) + p(4) + p(6) = 0.2 + 0.1 + 0.4 = 0.7,$$

因此

$$P(B | A) = \frac{P(A \text{ and } B)}{P(A)} = \frac{0.2}{0.7} = \frac{2}{7}.$$

1.4 贝叶斯定理

正如前面的练习所说明的，两个条件概率

$$P(A | B) \quad \text{and} \quad P(B | A)$$

一般来说并不相同。不过，它们仍然彼此相关。根据 Section 1.1 中的讨论，我们有两种不同的方法来确定概率 $P(A \text{ and } B)$ ：

$$P(A \text{ and } B) = P(A | B) \cdot P(B), \quad \text{and} \quad P(A \text{ and } B) = P(B | A) \cdot P(A).$$

我们可以把这两个表达式相等，并重新整理，得到：

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

这个关系被称为贝叶斯定理。

贝叶斯定理

对于一对事件 A 和 B ，如果 $P(A) \neq 0$ 且 $P(B) \neq 0$ ，那么

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

当我们知道一个条件概率，并想计算另一个条件概率时，贝叶斯定理特别有用。例如，根据上一节中不公平 dice 的例子，我们有：

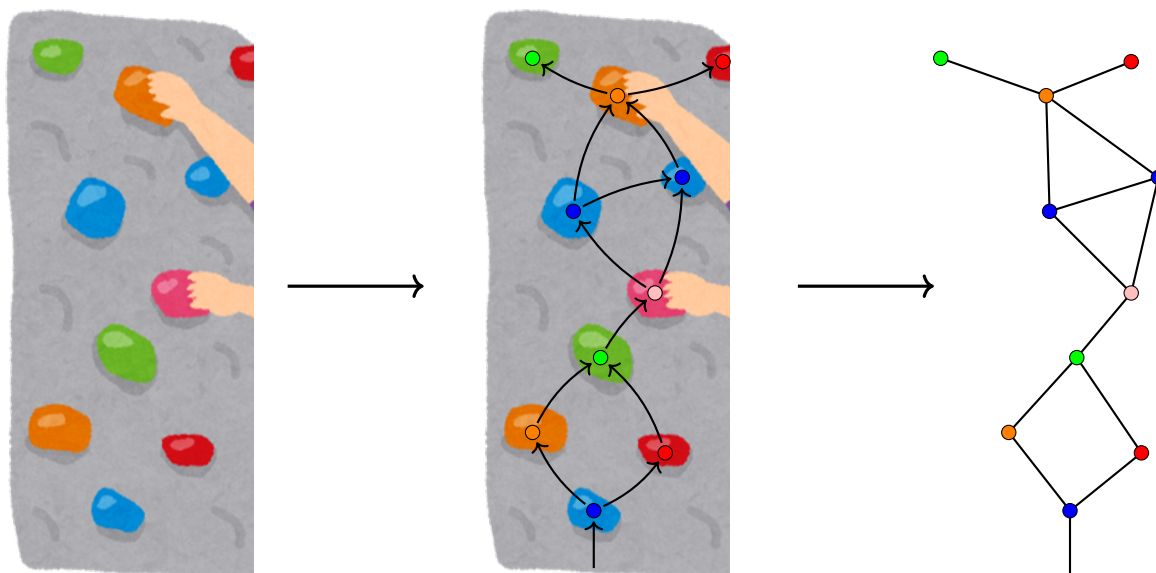
$$P(B | A) = \frac{P(A | B) \cdot P(B)}{P(A)} = \frac{0.5 \times 0.4}{0.7} = \frac{2}{7}.$$

2 决策树

当研究一系列可能相互影响的连续选择时，条件概率自然会出现。在小系统中，简单地画出所有可能发生事情的空间会非常有帮助。一种方法是使用决策树，这是一种根据多阶段过程的每一步进行分支的图。为了引入这个想法，我们先从一个相当复杂的例子开始。

2.1 一个引入例子：攀岩者

假设我们跟踪某人攀爬岩壁时手的移动。我们可以把这个情形理想化为一种“图”，它展示了彩色岩点网络之间的连接关系：



在这幅图中，随着攀岩者向上移动，他的手会从底部的蓝色岩点出发，向上沿着某条路径到达顶部的绿色或红色岩点。注意，在这个过程的某些阶段，攀岩者面临如何继续前进的选择。例如，在岩壁的最开始处，攀岩者可以选择从蓝色岩点到橙色岩点，也可以选择到红色岩点。在其他点，攀岩者不能选择该做什么，因为路径已经确定。例如，从第一个橙色岩点出发，必须沿着路径到达下一个绿色岩点。

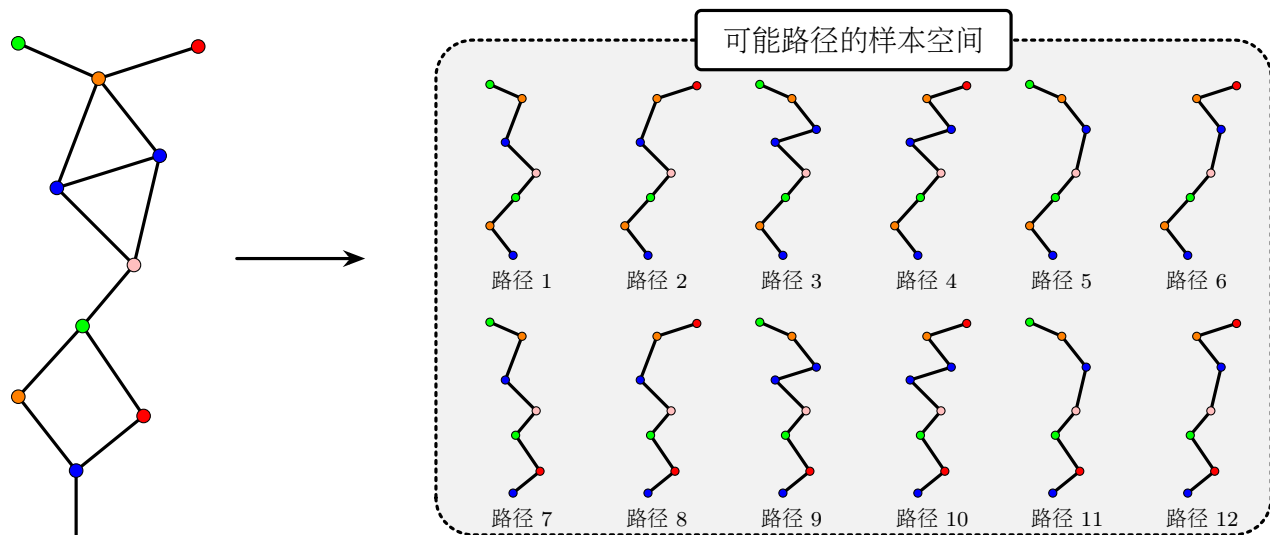
假设每次攀岩者做选择时，他有 50% 的概率选择左边的岩点，也有 50% 的概率选择右边的岩点。我们的核心问题是：

攀岩者走某一条特定路径的概率是多少？

假设有 12 条路径，一个最自然的猜测是应用古典概率的 Golden Formula:

$$P(\text{某一条特定路径}) = \frac{1}{\text{路径总数}}.$$

首先，我们应当描述样本空间。在这个例子中，样本空间是从底部到顶部的所有路径。花一点时间，你可以确信这面岩壁上实际有 12 条不同路径。它们如下图所示：



因此，在这个设置中，样本空间是集合

$$S = \{\text{路径 1}, \text{路径 2}, \text{路径 3}, \dots, \text{路径 12}\}.$$

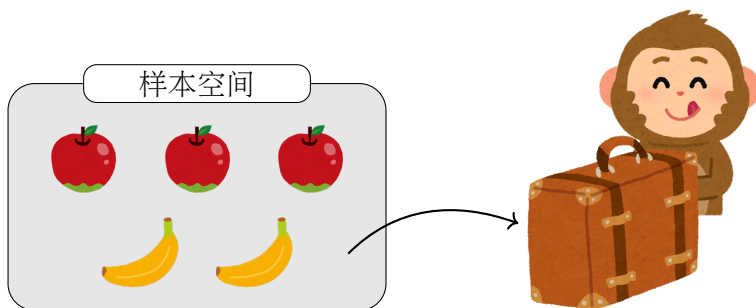
因此，尝试使用上面的公式来计算某条特定路径的概率似乎很合理。例如，对于路径1，我们可能猜测：

$$P(\text{路径 1}) \stackrel{?}{=} \frac{1}{\text{路径总数}} = \frac{1}{12}.$$

事实上，这是不正确的。稍后我们会正确地计算它，但现在我们只需要注意：在路径空间中，有些路径比其他路径更复杂。这些不同路径并不是等可能结果，所以使用上面的公式不是正确方法。我们很快会回到这一点。不过首先，我们先在一个更简单的设置中再次说明这个概念。

2.2 例子：丛林和小偷

假设你在丛林里，把你的包放在一旁无人看管。一只当地的猴子从树上爬下来，随机地翻你的包。碰巧，你包里有 5 个午餐水果：3 个苹果和 2 根香蕉。猴子开始随机偷水果，并且每找到一个就吃掉，然后再回到包里继续翻。在概率论中，这个过程叫做不放回抽取。



练习 1

- (a) 猴子第一次偷到苹果的概率是多少?
- (b) 猴子第一次偷到香蕉的概率是多少?

解答

总共有 5 个水果，所以可以进行简单计数：

- (a) $P(\text{第一次是苹果}) = \frac{3}{5}$.
- (b) $P(\text{第一次是香蕉}) = \frac{2}{5}$.

现在，我们关注一个连续问题：

他先偷一根香蕉，然后偷一个苹果的概率是多少？

显然，这里应该使用条件概率，因为第二次的结果明显受到第一次结果的影响。

为了精确表示顺序，令 B_1 表示“第一次抽到香蕉”，令 A_2 表示“第二次抽到苹果”。于是，“先香蕉后苹果”就是事件“ B_1 and A_2 ”。

我们可以用条件概率公式来计算。概率可以写成：

$$P(B_1 \text{ and } A_2) = P(A_2 | B_1) \cdot P(B_1).$$

一步一步来计算这个公式：

步骤 1: 我们之前已经看到 $P(B_1) = \frac{2}{5}$ 。

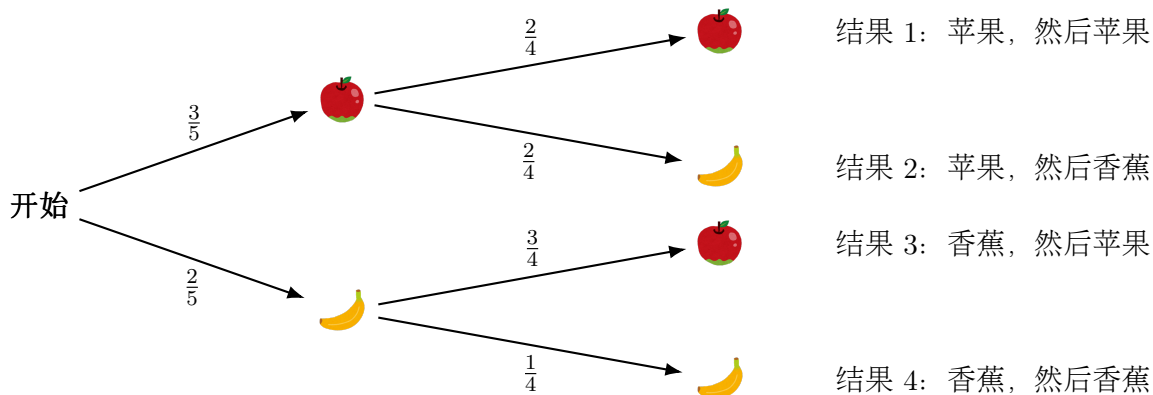
步骤 2: 现在计算 $P(A_2 | B_1)$ 。注意，如果一根香蕉被拿走并被猴子吃掉，那么第二次从包里取水果时，猴子面对的是一个不同的样本空间。第一次取水果时，有 5 个水果。然而，由于一根香蕉已经被拿走，现在只剩下 4 个水果：3 个苹果和 1 根香蕉。这个新的样本空间大小为 4，其中有 3 个苹果可供猴子选择。因此，条件概率为：

$$P(A_2 | B_1) = \frac{3}{4}.$$

步骤 3: 现在把这两个数相乘，得到先拿香蕉再拿苹果的概率：

$$P(B_1 \text{ and } A_2) = P(B_1) P(A_2 | B_1) = \frac{2}{5} \cdot \frac{3}{4} = \frac{3}{10}.$$

类似的推理可以应用到这只调皮猴子的其他可能情况。我们可以把整个过程整理成一棵树，如下所示：



如你所见，我们之前的计算，即

$$P(\text{香蕉然后苹果}) = P(B_1) \cdot P(A_2 | B_1) = \frac{2}{5} \cdot \frac{3}{4} = \frac{3}{10},$$

也可以通过检查上面树中对应结果3的分支来直观确认，因为这里我们只是把结果3对应分支上的数字相乘。事实上，我们刚刚看到了一个一般技巧的第一个例子：如果我们可以把一个多步骤实验的结果整理成一棵树，那么任意结果的概率都可以通过把该分支上的所有概率相乘得到。

用这个观察，我们可以快速计算树中的另外三个结果：

- 结果1: $P(A \text{ then } A) = \frac{3}{5} \cdot \frac{2}{4} = \frac{3}{10},$
- 结果2: $P(A \text{ then } B) = \frac{3}{5} \cdot \frac{2}{4} = \frac{3}{10},$
- 结果4: $P(B \text{ then } B) = \frac{2}{5} \cdot \frac{1}{4} = \frac{1}{10}.$

注意，这四个概率的和为

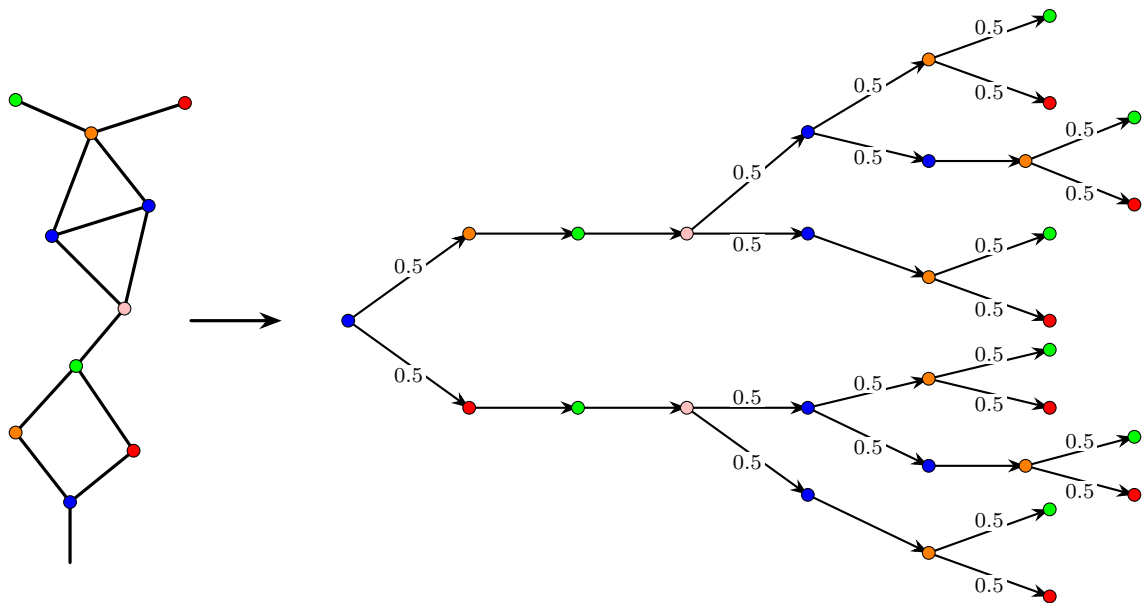
$$\frac{3}{10} + \frac{3}{10} + \frac{3}{10} + \frac{1}{10} = 1.$$

这等价于说，必然有其中一个结果会发生。

2.3 回到攀岩者

我们刚刚看到，决策树可以用来组织一个多步骤实验的所有可能结果。为了求某条特定路径的概率，我们只需要把写在该路径上的概率相乘。我们可以用这种技术来求攀岩者通过岩壁的每条路径的概率。

首先，我们为这个系统写出树，并在每个“选择点”标上假设的 50-50 概率。完整的树如下图所示。



上面的树被安排成路径按顺序排列：第一条路径对应路径1，第二条路径对应路径2，依此类推。你应该注意到，有些路径比其他路径更复杂，例如路径12看起来比路径9或10更简单。这是因为，有些上墙路径只需要做三次选择，而有些则需要做四次选择。事实上，任何包含“从粉色岩点向左移动”这一选择的路径，都会比那些改为向右移动的路径更复杂。无论如何，我们可以用决策树理论来计算攀岩者通过岩壁的任意路径的概率。和 klepto-monkey 的例子一样，在这个例子中，我们只需把该分支上写着的数相乘，就能得到总概率。结果总结如下。

各路径的概率

我们根据每条路径需要做的选择次数来区分：

- 路径 1, 2, 3, 4, 7, 8, 9 和 10 都需要做四次选择。在决策树上，对应这些路径的分支上有 4 个 0.5，所以它们的总概率为：

$$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{16}.$$

- 路径 5, 6, 11 和 12 只需要做三次选择。因此，走这些路径中每一条的概率都比其他路径更大：

$$\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}.$$

我们总是可以通过确认所有答案加起来等于 1 来检查计算是否正确。这里有：

$$8 \left(\frac{1}{16} \right) + 4 \left(\frac{1}{8} \right) = \frac{1}{2} + \frac{1}{2} = 1,$$

这是正确的。注意，这里我们还发现了另一件事：攀岩者走某条看起来更简单的路径的概率是 50%，不走这类路径的概率也是 50%。

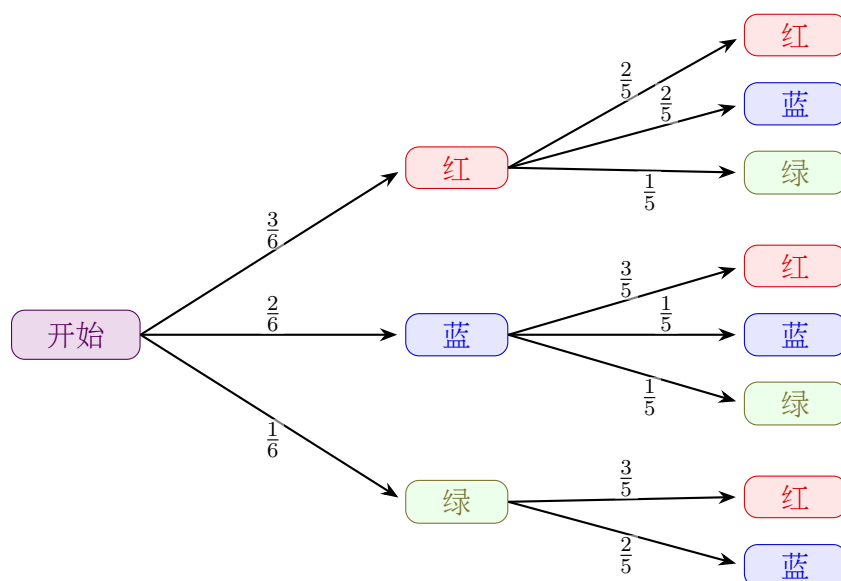
2.4 练习：不放回抽取

一个盒子里有 6 个球：3 个红球、2 个蓝球和 1 个绿球。你随机取出 2 个球，并且第一次取出的球不放回盒子中。

练习 2

- (a) 连续取出 2 个红球的概率是多少？
- (b) 先取出绿球，再取出蓝球的概率是多少？
- (c) 取出一个蓝球和一个红球，顺序不限的概率是多少？

在给出完整解答之前，我们先观察一下这棵树的样子。为了完整性，我们已经在每条箭头上标出了所有概率：



如你所见，在底部的分支上，不再有三个可能结果。为什么？因为盒子里只有 1 个绿球。如果我们第一次选中了它，那么盒子里就只剩下红球和蓝球。这棵树已经完整标出了所有概率和条件概率。我们不必直接用公式计算这些概率，而可以像 Section 2.2 中猴子偷水果的例子那样，用类似的直观推理。例如，要计算刚刚抽到一个红球之后又抽到一个红球的概率，我们考虑的情况是一个红球已经被拿走，因此样本空间现在缩小为剩下的 2 个红球，加上 2 个蓝球和 1 个绿球。由于在 5 个球中有 2 个红球可以选，条件概率 $P(\text{红} | \text{红})$ 就是 $\frac{2}{5}$ 。基于这一点，我们现在回答上面的问题。

解答

(a) 红球然后红球:

$$P(\text{红球然后红球}) = \frac{3}{6} \cdot \frac{2}{5} = \frac{6}{30} = \frac{1}{5}.$$

(b) 绿球然后蓝球:

$$P(\text{绿球然后蓝球}) = \frac{1}{6} \cdot \frac{2}{5} = \frac{2}{30} = \frac{1}{15}.$$

(c) 蓝球和红球表示红球然后蓝球, 或者蓝球然后红球:

$$P(\text{红球然后蓝球}) = \frac{3}{6} \cdot \frac{2}{5} = \frac{1}{5},$$

并且

$$P(\text{蓝球然后红球}) = \frac{2}{6} \cdot \frac{3}{5} = \frac{1}{5}.$$

这些是互斥结果, 因此:

$$P(\text{蓝球和红球}) = \frac{1}{5} + \frac{1}{5} = \frac{2}{5}.$$

3 著名例子

3.1 蒙提霍尔问题

假设你参加一个游戏节目。有三扇门。一扇门后面是一辆车, 另外两扇门后面是山羊。门后的物品是随机安排的, 所以你不知道哪扇门后面有什么。于是, 你选择门1, 希望车在后面。知道每扇门后面是什么的主持人打开另一扇门, 并展示一只山羊。假设主持人总是打开一扇不同的、后面有山羊的门, 并且总是给你换门的机会。现在你有一个选择:

坚持原来的门 或者 换到剩下那扇未打开的门。

你应该换吗?

令人惊讶的是, 在这些标准假设下, 答案是 yes: 换门更合理。坚持原门赢得车的概率是 $\frac{1}{3}$, 而换门赢得车的概率是 $\frac{2}{3}$ 。这很令人困惑, 所以我们用三种不同方式来解释它。

3.1.1 论证1: 转移概率

令

$$A = \{\text{你的第一次选择就是车所在的门}\}.$$

那么:

$$P(A) = \frac{1}{3},$$

因此:

$$P(\text{not } A) = \frac{2}{3}.$$

如果你的第一次选择是正确的，那么换门会输。如果你的第一次选择是错误的，那么换门会赢，因为主持人移除了另一只山羊。因此：

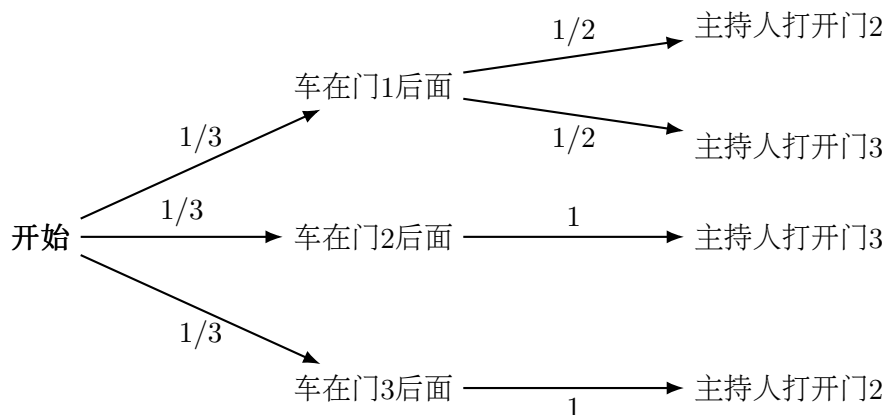
$$P(\text{换门获胜}) = P(\text{第一次选择是错误的}) = \frac{2}{3}.$$

3.1.2 论证2：把游戏变大

假设不是三扇门，而是有 1000 扇门，其中一扇门后面有车，另外 999 扇门后面有山羊。假设你和之前一样选择了门1。知道车实际在哪里的主持人打开了 998 扇门，并展示每扇门后面都是山羊。当只剩下两扇门时：你的门和另一扇门，你可能会开始怀疑自己的最初选择。如果有 1000 扇门，你最初选中的门正确的概率只有 $\frac{1}{1000}$ 。这个概率不会因为主持人打开了一些他已经知道是山羊的门就突然变大。相反，你的门仍然只有 $\frac{1}{1000}$ 的概率藏着车，而另一扇剩下的门极有可能藏着车。显然，换门会是更好的选择。

3.1.3 论证3：一棵决策树

假设你最初选择了门1，可能结果的空间如下所示。这里需要强调的是，主持人知道车实际在哪扇门后面。如果他刻意选择一扇会展示山羊的门，那么有时他有选择，有时没有。如果主持人有两扇山羊门可选，我们假设他等概率地在它们之间选择。



如果车恰好在门1后面，那么换门会输。根据上面的决策树，这种情况发生的概率是上面两条分支概率之和，结果是 $\frac{1}{3}$ 。如果车不在门1后面，那么它要么在门2后面，要么在门3后面。这种情况的概率是 $\frac{2}{3}$ 。因此，我们看到，换门会让你以 $\frac{2}{3}$ 的概率赢得汽车，而换门会让你以 $\frac{1}{3}$ 的概率失去汽车。

3.2 生日悖论

令 A 为事件“房间里至少两个人有相同生日”。在这种情况下，概率 $P(A)$ 会随着房间人数 n 的变化而发生巨大变化。例如，如果 $n = 2$ ，我们会预计这两个人生日相同的概率非常低。然而，如果 $n > 365$ ，那么忽略闰年时，因为一年只有 365 个可能生日，所以房间里必然有两个人生日相同。由此可见，当 n 很小时， $P(A)$ 也很小，而当 n 很大时， $P(A)$ 最终会变成 1。现在有一个问题：

当 n 等于多少时， $P(A)$ 会超过 50%？

换句话说：房间里需要有多少人，才能使至少两个人同生日这件事变得更可能发生？事实上，答案非常令人惊讶：只需要 23 个人！

这个惊人的事实被称为“生日悖论”。它是一个“真实性悖论”的例子，也就是说，一个数学结果如此令人意外，以至于它感觉上是错的。不过，正如我们现在要展示的，生日悖论可以由一些基本的概率应用推出。

3.2.1 解决生日悖论

为了解决生日悖论，我们将把问题反过来重述。我们不直接计算 $P(A)$ ，而是计算 $P(\text{not } A)$ ，也就是 A 的补事件的概率。回忆一下，补事件是概率论中对应于“不是”的概念。因此，这里的补事件 A^c 表示“房间里每个人的生日都不同”。

回忆一下，补事件的概率就是 $1 - P(A)$ 。所以，如果 $P(\text{not } A) < 50\%$ ，那么 $P(A) > 50\%$ ，因为

$$P(A) + P(\text{not } A) = 1.$$

我们假设每个人的生日选择是独立且等可能的，因此任意一天作为生日的概率都是 $1/365$ 。现在我们计算小 n 情况下的 $P(\text{not } A)$ ，看看过程如何运作。

当 $n = 2$ 时，我们要计算两个人生日不同的概率。显然，第一个人的生日可以自由选择。但是，为了让两个生日不同，第二个人的生日必须避开第一个人的生日。与第一个人的生日不同的日子有 364 天，所以如果把它看作随机事件，第二个人生日不同的概率是

$$P(\text{not } A) = \frac{\text{不同于第一个人生日的天数}}{\text{一年中的总天数}} = \frac{364}{365} \approx 0.997 = 99.7\%.$$

现在，假设我们向房间中加入第三个人，即 $n = 3$ 。我们刚刚确定，两个人生日不同的概率是 $\frac{364}{365}$ 。但我们还需要第三个人的生日不同于第一个人和第二个人的生日。下一个人的生日概率是以之前那些人的生日为条件的。要计算这种“and”语句的概率，我们需要把相关概率相乘：

$$\begin{aligned} P(\text{not } A) &= \frac{364}{365} \times \frac{\text{不同于第一个人和第二个人生日的天数}}{\text{一年中的总天数}} \\ &= \frac{364}{365} \cdot \frac{363}{365} \approx 0.992 = 99.2\%. \end{aligned}$$

我们可以对更大的 n 继续重复这个模式。每当 n 增加 1，我们就会继续“用掉”一些可供下一个人选择的生日。这意味着我们每一步都需要乘以越来越小的数。例如：

$$\begin{aligned} \text{当 } n = 4 : P(\text{not } A) &= \left(\frac{364}{365}\right) \left(\frac{363}{365}\right) \left(\frac{362}{365}\right) \approx 0.983 = 98.3\% \\ \text{当 } n = 5 : P(\text{not } A) &= \left(\frac{364}{365}\right) \left(\frac{363}{365}\right) \left(\frac{362}{365}\right) \left(\frac{361}{365}\right) \approx 0.973 = 97.3\% \\ \text{当 } n = 6 : P(\text{not } A) &= \left(\frac{364}{365}\right) \left(\frac{363}{365}\right) \left(\frac{362}{365}\right) \left(\frac{361}{365}\right) \left(\frac{360}{365}\right) \approx 0.960 = 96.0\% \end{aligned}$$

对于一般的 n ，我们得到公式：

$$P(\text{not } A) = \left(\frac{365}{365}\right) \left(\frac{364}{365}\right) \left(\frac{363}{365}\right) \cdots \left(\frac{365 - (n - 1)}{365}\right).$$

为了解决这个悖论，我们要找到使 $P(\text{not } A)$ 降到 50% 以下的最小 n ，因为这意味着 $P(A)$ 会超过 50%。如果我们把这个公式用于越来越大的 n ，就会看到当 $n = 23$ 时：

$$P(\text{not } A) = \left(\frac{365}{365}\right) \left(\frac{364}{365}\right) \left(\frac{363}{365}\right) \cdots \left(\frac{343}{365}\right) \approx 0.493 < 0.5.$$

相比之下，当 $n = 22$ 时， $P(\text{not } A) \approx 0.524 > 0.5$ 。所以，当 $n = 23$ 时， $P(A) > 50\%$ 。

3.3 医学检测

假设一个人口中有 1% 的人患有某种疾病。对于这种疾病，有一种医学检测，其中：

- 如果一个人患有这种疾病，那么检测有 99% 的概率检测出来；
- 如果一个人没有这种疾病，那么检测有 5% 的概率给出假阳性。

假设你接受了检测，并且检测结果为阳性。那么，你实际患有这种疾病的概率是多少？令人惊讶的是，答案大约是 16.7%。我们现在来展示原因。

首先，让我们用条件概率重写上面的条件。

令：

$$D = \{\text{此人患有疾病}\},$$

以及

$$T = \{\text{此人检测呈阳性}\}.$$

问题要求的是条件概率：

$$P(D | T),$$

也就是在检测呈阳性之后，实际患有疾病的概率。注意，这与反过来的概率不同： $P(T | D)$ 计算的是在一个人实际患病的条件下，检测呈阳性的概率。题目给出的条件告诉我们：

$$P(T | D) = 0.99, \quad P(D) = 0.01, \quad P(D^c) = 0.99, \quad \text{and} \quad P(T | D^c) = 0.05.$$

我们可以把这些信息与贝叶斯定理结合起来，计算缺失的条件概率：

$$P(D | T) = \frac{P(T | D)P(D)}{P(T)}.$$

为了做到这一点，我们需要先计算 $P(T)$ ，即整体上检测呈阳性的概率。

注意，有两种方式会检测呈阳性：

- 要么此人患病，并且正确检测为阳性；
- 要么此人没有患病，但错误检测为阳性。

这两种情况互斥，所以我们可以把事件 T 重新解释为一个“or”语句，并计算其概率：

$$P(T) = P(T | D)P(D) + P(T | D^c)P(D^c) = (0.99)(0.01) + (0.05)(0.99) = 0.0594.$$

把它代入贝叶斯定理，我们得到那个著名的 16.7%:

$$P(D | T) = \frac{P(T | D)P(D)}{P(T)} = \frac{0.99 \times 0.01}{0.0594} = \frac{1}{6} \approx 0.167.$$

为什么这感觉很惊人

当某人实际患有这种疾病时，这个检测非常擅长检测出疾病。然而，这种疾病很少见，而且假阳性的数量相对较多。这些假阳性降低了一次阳性结果的可靠性，从而导致了这个令人惊讶的结果。