

MAT220 第5讲：概率导论

目录

1 什么是概率?	2
1.1 不确定性	2
1.2 定义概率	2
1.3 古典概率	3
1.4 通过计数赋予概率	4
2 概率规则	6
2.1 简单事件、样本空间和事件	6
2.2 概率其实是集合论	7
3 复合概率	8
3.1 构造复合事件	8
3.2 补集: $\text{not } A$	9
3.3 交集: $A \text{ and } B$	10
3.4 条件概率	11
3.5 独立事件和依赖事件	13
3.6 独立事件的乘法规则	13
3.7 并集: $A \text{ or } B$	15
3.8 互斥事件	16

在第4讲中，我们介绍了概率论所需要的集合论和计数方法。今天，我们将用这些思想来从数学上描述概率。核心思想是：一个事件可以表示为可能结果的一个子集，而这个事件的概率是 0 到 1 之间的一个数。在古典概率论中，关于不确定性的问题可以转化为计数问题。在本讲中，我们将深入探讨这一思想。

1 什么是概率？

1.1 不确定性

在本课程最开始的几讲中，我们介绍了数据与统计的基本语言。我们看到，数据通常来自复杂的现实世界情境，然后被转化为可以用数学分析的形式。今天，我们将处理另一种复杂性：不确定性。

考虑下面两个命题：

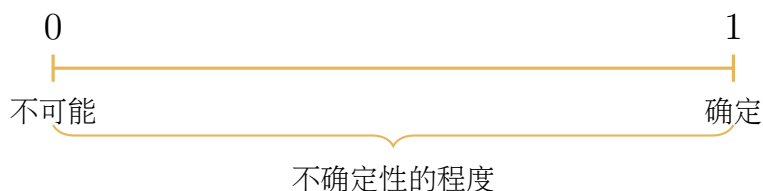
$$2 + 2 = 4 \quad \text{and} \quad 2 + 2 = 5.$$

你大概已经知道，第一个命题是真的，第二个命题是假的。这不仅显然，而且是明确无误的：这些命题没有其他可能的情况。相比之下，考虑下面两个命题：

“Dr. O’Connell 的中间名是 John”，以及“Dr. O’Connell 的中间名是 Jennifer”。

可以看到，这些非数学命题就不再那么清楚了：如果你碰巧不认识 Dr. O’Connell，那么你就会不确定他的中间名是什么。然而，即使你不知道我的中间名，你仍然可以想象，对于一个英格兰人来说，中间名叫 Jennifer 会比较少见，而中间名叫 John 则更合理。因此，虽然你不能明确指出这些命题的确切真假，但你仍然可以比较它们的可信程度。

概率论试图用数字来刻画这种不确定性。现实中，世界充满了并非纯粹逻辑性的未知事物，而概率论给了我们一种量化这些信念的方法，使它们可以相互比较。如果我们把数字 0 解释为“绝对不可能”，把数字 1 解释为“绝对确定”，那么我们所有不确定的命题都会位于二者之间的某个位置：



基本思想

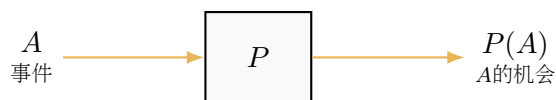
概率是不确定性的数学。它给我们一种给事件赋予数字的方法，使我们可以比较这些事件有多可能发生。

需要注意的是，概率**不是**真理的光谱。¹相反，它是确定性的光谱。因此，概率值 0.5 并不意味着某件事“半真半假”，而是意味着我们给这个命题为真的可能性赋予 50% 的机会。

1.2 定义概率

从数学上说，概率是一个给事件赋予数字的函数。如果字母 A 表示某件可能发生也可能不发生的事情，那么我们把 A 发生的概率写作 $P(A)$ 。在这个意义上，我们可以把字母 P 看作一台小机器：它吃进去现实世界中的事件，然后吐出一个表示其确定程度的数值。

¹给感兴趣的人：实际上有一种逻辑理论确实研究这种事情，叫作“模糊逻辑”。但这**不是**概率。



一般来说，我们把这些“可能发生的事情”称为事件，并用 A, B, C 等大写字母表示。形式上，赋予事件的概率可以总结如下。

概率值

对于任意事件 A ，它的概率满足

$$0 \leq P(A) \leq 1.$$

这里， $P(A) = 0$ 表示 A 是不可能的，而 $P(A) = 1$ 表示 A 是确定的。

由于概率会映射到 0 和 1 之间的某个位置，我们可以直接比较不同事件的不确定程度。假设有两个事件 A 和 B ，并且我们知道它们的概率 $P(A)$ 和 $P(B)$ 。如果 $P(A) < P(B)$ ，那么事件 B 比事件 A 更可能发生。例如，如果

$$P(A) = 0.2 \quad \text{and} \quad P(B) = 0.7,$$

那么 B 比 A 更可能发生，因为 $P(B)$ 比 $P(A)$ 更接近绝对确定性 1。

1.2.1 表示概率

由于任意概率都是 0 到 1 之间的数，所以我们可以用几种等价的方式写出它。例如，如果事件 A 的概率是 50%，我们可以用下面任意一种方式表示：

$$P(A) = \frac{1}{2} = 0.5 = 50\% = \frac{250}{500}.$$

这里， $\frac{1}{2}$ 是最简分数，0.5 是小数，50% 是百分数，而 $\frac{250}{500}$ 是未约分的分数。在最后一种情况下，也许我们正在处理一个有 500 种可能性的系统，其中 250 种可能结果是有利的。未约分的分数 $\frac{250}{500}$ 告诉了我们这些信息，而最简分数 $\frac{1}{2}$ 则没有告诉我们这些具体信息。

1.3 古典概率

在本讲中，我们主要关注古典概率。这是指可能结果有限，并且每个结果同样可能发生的情形。

定义：古典概率

当可能结果有限，并且每个结果同样可能发生时，这个概率问题就是**古典概率**问题。

例子包括：

- 掷一枚公平的骰子，
- 抛一枚公平的硬币，
- 从充分洗混的一副牌中发牌，
- 选择判断题的答案。

在这种情形下，概率通常可以通过计数来计算。

1.3.1 关于随机性的说明

真正的随机性是一个很深的哲学问题。有些人认为世界是完全决定论的，意思是所有事情都由过去的事件和物理定律决定。也有人认为确实存在真正随机的事情，例如量子力学中的观测，但即使在那里也仍有争论。无论如何，对我们的目的来说，真正的随机性是否存在并不重要。在本课程中，我们用概率来量化不确定性。因此，即使事物在现实中并不真正随机，它们仍然可能在我们看来是随机的。

本课程中的随机性

在本课程中，如果我们在实验进行之前不能确定地预测某个结果，那么这个结果就是随机的。

这种实用解释已经足够用于本课程将要进行的概率计算。

1.4 通过计数赋予概率

在古典概率中，我们可以通过把“有利情况”的数量与所有可能性的总数相比较，来计算单个事件的概率。对于任意事件，我们可以用下面的公式赋予概率：

$$\text{事件的概率} = \frac{\text{对该事件有利的结果数量}}{\text{结果总数}}.$$

换句话说，我们数出会使该事件为真的结果数量，然后除以可能结果的总数。在这个分数中，分子总是小于或等于分母，所以取比值时总会得到 0 到 1 之间的值。

古典概率的黄金公式

如果所有结果同样可能发生，那么

$$P(A) = \frac{\text{对 } A \text{ 有利的结果数量}}{\text{结果总数}}.$$

现在我们将展示如何使用这个公式来计算一些基本概率。

1.4.1 单个结果的概率

任意单个结果的概率都可以用这个公式确定。例如：

- 对于一枚公平的硬币，硬币落在正面的概率是 $\frac{1}{2}$ ，因为只有两种可能性，其中一种是正面。
- 掷一枚公平的骰子时，掷出 5 的概率是 $\frac{1}{6}$ ，因为骰子上只有一个 5，总共有 6 个数字。
- 从洗牌的一副牌中随机抽一张牌时，抽到黑桃 A 的概率是 $\frac{1}{52}$ ，因为一副牌有 52 张，而其中只有一张是黑桃 A。

一般来说，对于古典概率中的任意系统，上面的公式告诉我们：

$$\text{任意单个结果的概率} = \frac{1}{\text{结果总数}}.$$

1.4.2 更复杂事件的概率

我们也可以用黄金公式来计算更复杂的概率。如前所述，这可以通过查看所有可能性的空间，并数出所有使该事件为真的结果来完成。作为例子，假设我们正在掷一枚公平的骰子。令事件 A 为：

$A = \text{“掷出偶数”}$ ，

并令事件 B 为：

$B = \text{“掷出的数小于 5”}$ 。

为了确定这些概率，我们需要计算每个事件的“有利结果”数量，然后除以 6，因为骰子有 6 个面。对于事件 A ，我们有：

$$P(A) = \frac{\text{偶数有多少个}}{\text{总共有多少个数}} = \frac{3}{6} = \frac{1}{2},$$

同样，对于 B ，概率是：

$$P(B) = \frac{\text{小于 5 的数有多少个}}{\text{总共有多少个数}} = \frac{4}{6} = \frac{2}{3}.$$

1.4.3 例：玩具柜

假设你过分热情的表亲有一个柜子，里面有 58 个不同的玩具，他随机挑一个给你看。在他这样做之前，你不知道他会选哪一个，所以从你的角度来看，他的选择本质上是随机的。观察下面的玩具柜：



令事件 A 为：“被选中的玩具是一辆蓝色汽车”。通过查看上面的图片，我们看到柜子里有两辆蓝色汽车。因此，根据黄金公式，

$$P(A) = \frac{2}{58}.$$

令事件 B 为：“被选中的玩具来自第四排”。在上图中，我们数出第四排包含 8 个玩具，所以事件 B 包含 8 个结果。因此：

$$P(B) = \frac{8}{58}.$$

两种情况下的计算是一样的：数出有利玩具，然后除以玩具总数。

练习：判断题答案序列

假设三道作业题都是判断题。为了随机选择正确答案，Dr. O'Connell 写下所有可能的真/假序列，并从帽子中抽取一个。连续 3 个 T/F 的可能序列共有 8 个：

TTT, TTF, TFT, FTT, TFF, FTF, FFT, FFF

- (a) 三个答案全是 False 的概率是多少？
- (b) 恰好两个答案是 False 的概率是多少？
- (c) 如果一名学生盲猜所有三个答案，那么他三题全答对的概率是多少？

解答

这三个问题都可以用前面提到的黄金公式回答。首先注意，可能的真/假序列共有 8 个。

- (a) 有利结果只有一个，即 FFF。因此

$$P(\text{三个全是 False}) = \frac{1}{8}.$$

- (b) 有利结果是 TFF、FTF 和 FFT。因此

$$P(\text{恰好两个是 False}) = \frac{3}{8}.$$

- (c) 正确序列恰好只有一个，所以

$$P(\text{三题全答对}) = \frac{1}{8}.$$

2 概率规则

2.1 简单事件、样本空间和事件

现在我们更仔细地整理概率的词汇。最重要的对象是简单事件、样本空间和事件。

重要定义

- (1) **简单事件**由实验的恰好一个结果组成。
- (2) **样本空间**通常写作 S ，是所有可能结果的集合。
- (3) **事件**是样本空间的任意子集，意思是它可以包含一个结果、多个结果，或者完全不包含结果。
- (4) 事件 A 的概率写作 $P(A)$ 。这是 0 到 1 之间的一个数。
- (5) 所有简单事件的概率之和总是 1。

为了看到这些定义如何使用，让我们回到玩具柜的例子。在这个例子中，样本空间是柜子里所有 58 个玩具的集合：

$$S = \{\text{柜子里的所有玩具}\}.$$

在这个例子中，一个简单事件就是选择一个单独的玩具，例如“选中这个特定玩具”。从数学上说，这就是 S 的任意一个大小为一的子集。一般事件则是 S 的任意子集，这些子集表示更一般的陈述，例如“选中第四排的一个玩具”或“选中一辆蓝色汽车”。

再次强调：事件就是样本空间的子集。例如，在上面的玩具柜例子中，我们考虑了事件 A ，意思是“选中一辆蓝色汽车”，并且求得它的概率为 $\frac{2}{58}$ 。事件 A 的数学定义就是包含两个对象的集合：从柜子中取出的那两辆蓝色汽车。

2.2 概率其实是集合论

我们再考虑一个例子。假设 Dr. O'Connell 的班上有 11 名学生。他让他们排成两排：前排 6 名学生，后排 5 名学生：



为了讨论方便，Dr. O'Connell 随机从前排选一个人的概率是多少？假设每个人被选中的可能性相同，根据我们的黄金公式：

$$P(A) = \frac{\text{前排学生人数}}{\text{学生总人数}} = \frac{6}{11}.$$

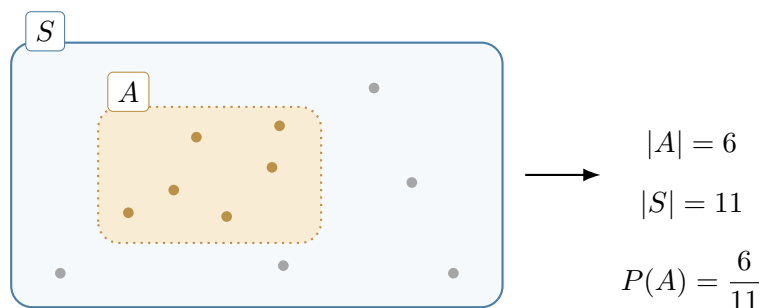
用前面的术语来说，样本空间 S 对应于所有学生的集合，而事件 A 是由前排学生组成的子集：

$$A = \left\{ \begin{array}{c} \text{前排学生 1} \\ \text{前排学生 2} \\ \text{前排学生 3} \\ \text{前排学生 4} \\ \text{前排学生 5} \\ \text{前排学生 6} \end{array} \right\}$$
$$S = \left\{ \begin{array}{c} \text{前排学生 1} \\ \text{前排学生 2} \\ \text{前排学生 3} \\ \text{前排学生 4} \\ \text{前排学生 5} \\ \text{前排学生 6} \\ \text{后排学生 1} \\ \text{后排学生 2} \\ \text{后排学生 3} \\ \text{后排学生 4} \\ \text{后排学生 5} \end{array} \right\}$$

快速数一下可知， $|A| = 6$ 且 $|S| = 11$ 。取这两个数的比就得到概率 $P(A)$ 。因此，用集合论语言来说，我们的黄金公式现在可以写成：

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

这说明，要计算事件 A 的概率，我们数出子集 A 的大小，然后除以样本空间 S 的大小。更抽象地说，我们正在进行如下图所示的计数。



为了以后引用，我们重新表述这个公式。

古典概率黄金公式的集合形式

如果所有结果同样可能发生，那么事件 A 的概率可以由

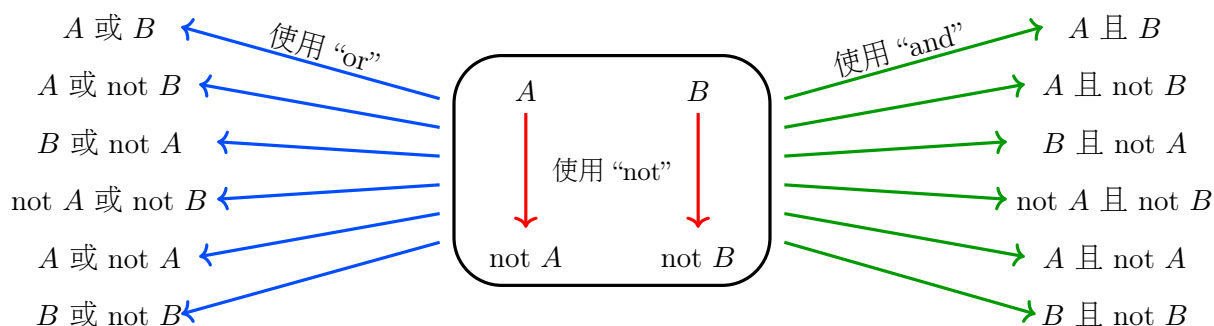
$$P(A) = \frac{|A|}{|S|}$$

确定，也就是说，用 A 的基数除以样本空间 S 的基数。

3 复合概率

3.1 构造复合事件

假设我们有两个事件 A 和 B 。根据目前的讨论，我们知道如何确定概率 $P(A)$ 和 $P(B)$ ：只需数出对应于 A 或 B 的结果数量，然后除以样本空间的大小。事实上，我们可以通过三个逻辑构件从 A 和 B 构造许多其他事件，即：*not*、*and* 和 *or*。例如，把这三个词用于 A 和 B ，可以创造出许多其他陈述，而它们本身也都是事件：



这个图看起来很像我们在第4讲讨论集合论时见过的图。事实上，它就是同一个图。*not*、*and* 和 *or* 这三个操作可以直接翻译成集合论如下：

语言	集合运算	含义
not A	A^c	A 的补集
A and B	$A \cap B$	A 与 B 的交集
A or B	$A \cup B$	A 与 B 的并集

根据前一节，我们可以通过数出任意事件对应子集的大小，再除以样本空间的大小，来计算该事件的概率。对于上面的三个陈述，这给出了三个新的公式：

$$P(\text{not } A) = \frac{|A^c|}{|S|},$$

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

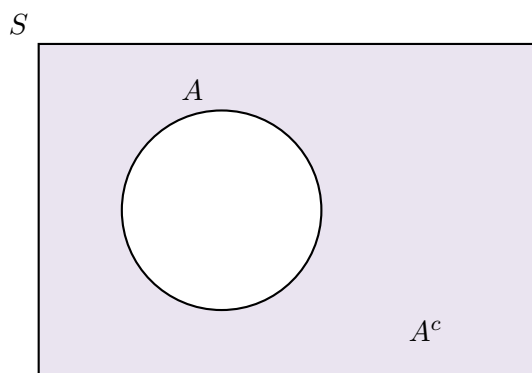
以及

$$P(A \text{ or } B) = \frac{|A \cup B|}{|S|}.$$

因此，如果我们能算出这些集合的大小，就可以建立公式来计算相应的复合概率。现在我们逐一解释。

3.2 补集：not A

补集 A^c 指样本空间中不属于子集 A 的所有东西。在下图中，补集是阴影区域。



根据古典概率公式，要计算 not- A 的概率，我们应该数出补集 A^c 的大小：

$$P(\text{not } A) = \frac{|A^c|}{|S|}.$$

由于 A 和 A^c 合在一起覆盖整个样本空间，我们有

$$|A| + |A^c| = |S|.$$

因此

$$|A^c| = |S| - |A|.$$

把它代入概率公式，得到

$$P(A^c) = \frac{|A^c|}{|S|} = \frac{|S| - |A|}{|S|} = \frac{|S|}{|S|} - \frac{|A|}{|S|} = 1 - P(A).$$

补集规则

对于任意事件 A ,

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

例。 在真/假序列的例子中，令 E 为事件“选中 FFF”。由于有 8 个同样可能的序列，

$$P(E) = \frac{1}{8}.$$

补集 E^c 是事件“没有选中 FFF”。有利结果有 7 个，也就是除了 FFF 之外的所有结果。因此，

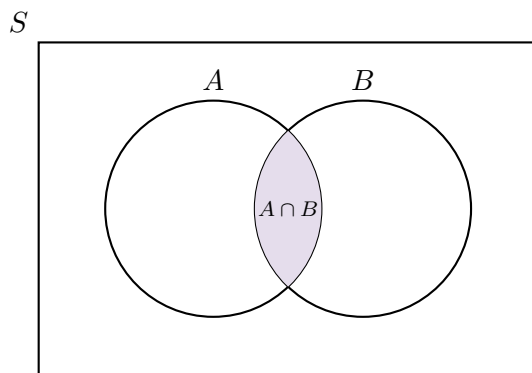
$$P(E^c) = \frac{7}{8}.$$

等价地，使用补集规则，

$$P(E^c) = 1 - P(E) = 1 - \frac{1}{8} = \frac{7}{8}.$$

3.3 交集：A and B

A 与 B 的交集写作 $A \cap B$ ，它是同时属于 A 和 B 的结果集合。视觉上，它就是两个集合的重叠部分：



根据黄金公式，复合事件 A and B 的概率应该涉及它们交集的大小：

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

遗憾的是，仅仅根据 $|A|$ 和 $|B|$ ，并没有一个通用公式可以确定 $A \cap B$ 的大小。两个大集合可能重叠很多，可能只重叠一点，也可能完全不重叠。因此，要计算 $P(A \text{ and } B)$ ，我们需要知道这两个事件彼此之间的关系。

3.4 条件概率

假设我们尝试对 $P(A \text{ and } B)$ 的公式做一些代数变形。我们聪明地人为插入一个 1，把它伪装成 $\frac{|B|}{|B|}$ 的形式。进行一些基本代数运算，我们得到：

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

这样，我们成功把这个概率表示为两个其他比值的乘积。第二个分数很熟悉：它就是 $P(B)$ 的定义。相比之下，第一个比值是新的：它是一种特殊概率，称为条件概率。

定义：条件概率

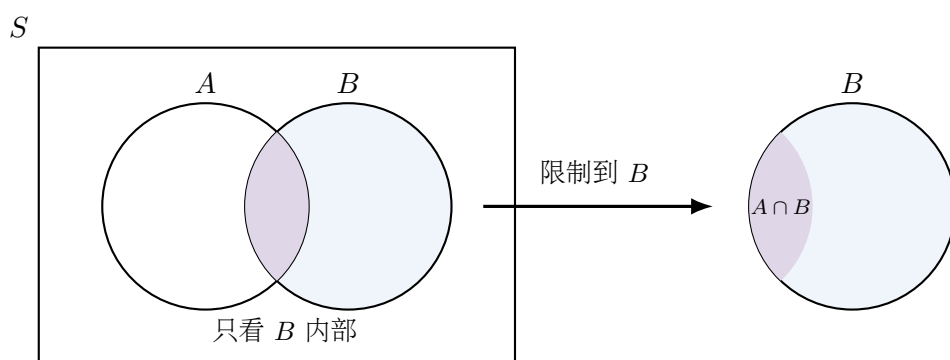
条件概率 $P(A | B)$ 是在假设 B 已经发生的情况下， A 发生的概率。在古典概率中，

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

前提是 $|B| \neq 0$ 。

条件概率这个概念乍看起来会令人困惑。因此，我们现在用集合论来解释发生了什么。

如前所述，没有一个完美公式可以用来计算交集 $A \cap B$ 的大小。弄清这个大小的一种方法是只看集合 B 内部，然后数出其中所有属于 A 的情况。这就像把整个样本空间拿来，然后选择性地忽略所有不已经在 B 中的东西：



如果我们尝试基于这个新设置来计算概率，就会注意到样本空间不再是 S ，而是子集 B 。在这个意义上，由于样本空间现在是 B ，如果我们试图在这个受限情境中计算 A 发生的概率，就需要在黄金公式的分母中使用 B ：

$$P(A, \text{ 在 } B \text{ 内}) = \frac{\text{既属于 } A \text{ 又属于 } B \text{ 的元素数量}}{B \text{ 的大小}}.$$

任何既属于 A 又属于 B 的元素都恰好落在交集 $A \cap B$ 中。因此，上面的概率使我们可以恢复之前的条件概率公式：

$$P(A | B) = P(A, \text{ 在 } B \text{ 内}) = \frac{\text{既属于 } A \text{ 又属于 } B \text{ 的元素数量}}{B \text{ 的大小}} = \frac{|A \cap B|}{|B|}.$$

作为例子，假设我们掷一枚公平的骰子。令

$$A = \{\text{掷出偶数}\} \quad \text{and} \quad B = \{\text{掷出小于 4 的数}\}.$$

写成子集形式，事件 A 和 B 是：

$$A = \{2, 4, 6\}, \quad \text{and} \quad B = \{1, 2, 3\}.$$

于是，你把骰子扔到空中。想象一下，在骰子落地之前，上帝从天而降，在你耳边低语：“结果会小于 4”。这意味着在那一刻，你确定知道不可能掷出 4、5 或 6。在这种情况下，如果你要评估事件 A 发生的机会并计算新的概率，你会把可能性空间从

$$\{1, 2, 3, 4, 5, 6\}$$

缩小到较小的集合

$$\{1, 2, 3\}.$$

在这个较小集合中，只有一个结果是偶数，即 2。因此

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

使用集合记号，

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

所以

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

把这些合在一起，我们看到

$$P(B) = \frac{3}{6} = \frac{1}{2}, \quad P(A \text{ and } B) = \frac{1}{6}, \quad P(A | B) = \frac{1}{3}.$$

有趣的是，刚好有：

$$\frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3}.$$

但这不是巧合：它体现了我们前面提到的一般公式。

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

一般乘法规则

当相关的条件概率有定义时，对于任意事件 A 和 B ，

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

等价地，

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

3.5 独立事件和依赖事件

在某些情况下， $P(A \text{ and } B)$ 的公式可以大大简化。为了正确解释如何简化，我们首先需要区分独立事件和依赖事件。

独立事件和依赖事件

令 A 和 B 为事件。

- 如果知道其中一个事件发生不会改变另一个事件的概率，那么 A 和 B 是**独立的**。
- 如果知道其中一个事件发生会改变另一个事件的概率，那么 A 和 B 是**依赖的**。

回忆一下，概率是不确定性的科学。因此，依赖事件背后的想法是：知道事件 B 已经发生，可能会影响事件 A 的概率。当关于 B 的信息以某种方式与事件 A 有关时，这种情况就可能发生。

在前面掷骰子的例子中，事件 A 是“掷出偶数”，事件 B 是“掷出小于 4 的数”，它们是依赖事件，因为知道 B 会发生，实际上把我们的概率从 $\frac{1}{2}$ 改变为 $\frac{1}{3}$ 。换句话说，关于一个事件的信息影响了我们对另一个事件可能性的认识。

然而，我们也可以想象某些情形，在这些情形中，知道一条信息并不会给我们理由改变概率判断。例如，令事件 C 为“掷出大于 2 的数”。那么作为子集，

$$C = \{3, 4, 5, 6\}.$$

如果由于某种原因，我们绝对知道骰子的结果会大于 2，那么我们可以把样本空间从

$$\{1, 2, 3, 4, 5, 6\}$$

限制到子集

$$\{3, 4, 5, 6\}.$$

由于这个受限集合由两个奇数和两个偶数组成，掷出偶数的概率仍然是 $\frac{1}{2}$ 。这与 $P(A)$ 相同，所以知道 C 会发生并不会改变我们对这个系统的确定程度。因此，事件 A 和 C 是独立的。

用符号表示，当

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

时， A 和 B 是独立的；等价地，

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

3.6 独立事件的乘法规则

如果 A 和 B 独立，那么可以用 $P(A | B) = P(A)$ 来大幅简化 $P(A \text{ and } B)$ 的公式。我们可以进行代入，得到下面的乘法规则。

独立事件的乘法规则

如果 A 和 B 是独立事件，那么

$$P(A \text{ and } B) = P(A)P(B).$$

例：两次抛硬币。 假设我们抛一枚公平硬币两次。

令 A 为事件“第一次抛出反面”，令 B 为事件“第二次抛出反面”。这些事件是独立的，因为第一次的结果不会影响第二次的结果。

因此，

$$P(A \text{ and } B) = P(A)P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

等价地，样本空间是

$$S = \{HH, HT, TH, TT\},$$

而只有 TT 是有利结果。

例：两副不同的牌。 假设我们洗混两副不同的牌。

令 A 为事件“第1副牌最上面的牌是 A”，令 B 为事件“第2副牌最上面的牌是 A”。这些事件是独立的，并且每个事件的概率都是

$$\frac{4}{52} = \frac{1}{13}.$$

所以

$$P(A \text{ and } B) = \left(\frac{1}{13}\right) \left(\frac{1}{13}\right) = \frac{1}{169}.$$

等价地，最上面两张牌的有序对共有 $52 \times 52 = 2704$ 种可能，其中 $4 \times 4 = 16$ 种由两张 A 组成。因此

$$\frac{16}{2704} = \frac{1}{169}.$$

下面是一个依赖事件的例子。

例：从一副牌中抽出两张 A。 假设我们洗混一副 52 张的牌。令

$$A = \{\text{第一张牌是 A}\},$$

并且

$$B = \{\text{第二张牌是 A}\}.$$

这些事件是依赖的，因为如果第一张牌是 A，那么抽第二张牌时，剩下的 51 张牌中只有 3 张 A。我们计算：

$$P(A) = \frac{4}{52} = \frac{1}{13},$$

并且

$$P(B | A) = \frac{3}{51}.$$

因此，

$$P(A \text{ and } B) = P(A)P(B | A) = \frac{1}{13} \cdot \frac{3}{51} = \frac{1}{221}.$$

如何计算 $P(A \text{ and } B)$

步骤 1: 判断事件是否独立。如果 $P(A) = P(A | B)$, 那么 A 和 B 是独立的。

步骤 2: 如果 A 和 B 独立, 使用

$$P(A \text{ and } B) = P(A)P(B).$$

步骤 3: 对于任意事件, 无论独立还是依赖, 都可以使用

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

或者等价地,

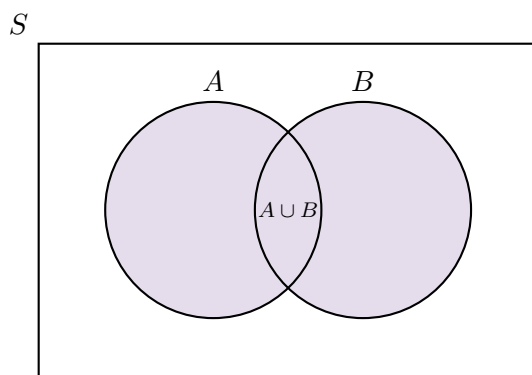
$$P(A \text{ and } B) = P(B)P(A | B).$$

3.7 并集: $A \text{ or } B$

根据 Section 3.1, 复合事件“ $A \text{ or } B$ ”对应于集合论中的并集 $A \cup B$ 。因此, 为了计算概率 $P(A \text{ or } B)$, 我们可以数出这个子集的大小, 然后除以样本空间的大小:

$$P(A \text{ or } B) = \frac{|A \cup B|}{|S|}.$$

A 和 B 的并集写作 $A \cup B$, 它是属于 A 、属于 B 或同时属于二者的结果集合。



与前面讨论的交集一样, 一般来说, 我们不能仅根据 A 和 B 的大小推出并集 $A \cup B$ 的大小。事实上, 我们在上一讲注意到, 很容易天真地写出:

$$|A \cup B| = |A| + |B|.$$

然而, 这一般是错误的, 因为重叠部分 $A \cap B$ 会被意外地数两次。例如, 如果

$$A = \{1, 2, 3\}, \quad \text{and} \quad B = \{2, 3, 4\},$$

那么它们的并集是

$$A \cup B = \{1, 2, 3, 4\},$$

它的基数等于 4。然而,

$$|A| + |B| = 3 + 3 = 6 \neq 4.$$

问题在于，交集 $\{2, 3\}$ 的两份副本隐藏在和 $3 + 3 = 6$ 中。

一般来说，为了解决这个问题，我们可以减去一份交集：

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

我们可以利用这个观察，推导出用 $P(A)$ 、 $P(B)$ 和 $P(A \text{ and } B)$ 表示 $P(A \text{ or } B)$ 的一般公式：

$$\begin{aligned} P(A \text{ or } B) &= \frac{|A \cup B|}{|S|} \\ &= \frac{|A| + |B| - |A \cap B|}{|S|} \\ &= \frac{|A|}{|S|} + \frac{|B|}{|S|} - \frac{|A \cap B|}{|S|} \\ &= P(A) + P(B) - P(A \text{ and } B). \end{aligned}$$

这称为加法规则，值得单独放入一个框中。

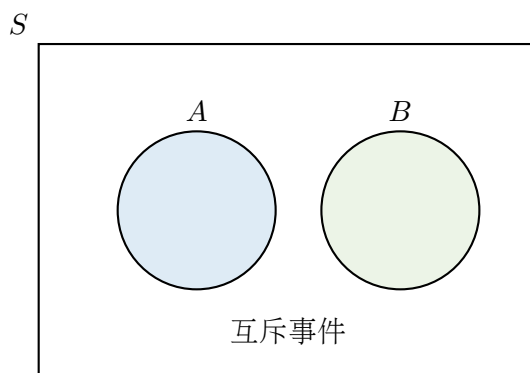
加法规则

对于任意事件 A 和 B ,

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B).$$

3.8 互斥事件

有时两个事件不能同时发生。例如，一次抛硬币不可能同时得到正面和反面。这样的事件称为互斥事件。视觉上，互斥事件没有重叠：



定义：互斥事件

两个事件 A 和 B 如果不能同时发生，则称为互斥。等价地， $A \cap B = \emptyset$ ，这意味着

$$P(A \text{ and } B) = \frac{|A \cap B|}{|S|} = \frac{0}{|S|} = 0.$$

利用这个观察，我们现在注意到，在 A or B 的概率公式中，对于互斥事件，第三项会消失：

$$\begin{aligned}
 P(A \text{ or } B) &= \frac{|A \cup B|}{|S|} \\
 &= \frac{|A| + |B| - |A \cap B|}{|S|} \\
 &= \frac{|A|}{|S|} + \frac{|B|}{|S|} - \frac{|A \cap B|}{|S|} \\
 &= P(A) + P(B) - P(A \text{ and } B) \\
 &= P(A) + P(B).
 \end{aligned}$$

这意味着我们的公式会化简为 A 和 B 的概率的简单相加。这称为互斥事件的加法规则。

互斥事件的加法规则

如果 A 和 B 互斥，那么

$$P(A \text{ or } B) = P(A) + P(B).$$

作为例子，假设我们洗混一副标准的 52 张牌。定义事件：

$A = \{\text{最上面的牌是 A}\},$

$B = \{\text{最上面的牌是黑桃}\},$

以及

$C = \{\text{最上面的牌是 K}\}.$

事件 A 和 B 不是互斥的，因为黑桃 A 既是 A 又是黑桃。然而， A 和 C 是互斥的，因为最上面的牌不可能同时是 A 和 K。

练习：A、黑桃和 K

计算下列概率。

(a) $P(A \text{ or } B)$ 。

(b) $P(A \text{ or } C)$ 。

解答

(a) 我们有

$$P(A) = \frac{4}{52}, \quad P(B) = \frac{13}{52}, \quad P(A \text{ and } B) = \frac{1}{52}.$$

因此，

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) = \frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}.$$

(b) 由于 A 和 C 互斥，

$$P(A \text{ or } C) = P(A) + P(C).$$

因此，

$$P(A \text{ or } C) = \frac{4}{52} + \frac{4}{52} = \frac{8}{52} = \frac{2}{13}.$$

总结：复合概率

本讲中的主要复合事件由 *not*、*and* 和 *or* 这几个词构造而成。在集合记号中，它们对应为

$$\text{not } A \longleftrightarrow A^c, \quad A \text{ and } B \longleftrightarrow A \cap B, \quad A \text{ or } B \longleftrightarrow A \cup B.$$

对于补集，使用

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

对于交集，使用乘法规则

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

或者等价地，

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

如果事件独立，则化简为

$$P(A \text{ and } B) = P(A)P(B).$$

对于并集，使用加法规则

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B).$$

如果事件互斥，那么 $P(A \text{ and } B) = 0$ ，所以它化简为

$$P(A \text{ or } B) = P(A) + P(B).$$