

MAT120: 论证 第10讲讲义

概率论导论

在本课程的上一部分中，我们学习了建模，也就是说，我们用方程来描述变量之间的关系。在本讲中，我们开始一个新主题：概率。概率是不确定性的数学。它为我们提供了一种讨论事件发生可能性的方式，也将成为接下来统计单元的基础。

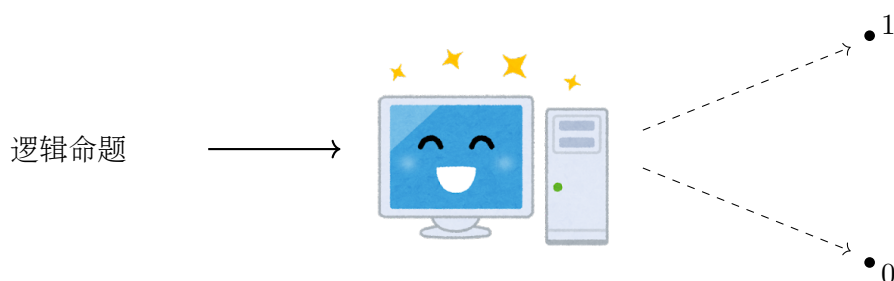
1 什么是概率？

1.1 逻辑与不确定性

在本课程最开始的几讲中，我们介绍了逻辑。逻辑是一种“形式语言”，我们可以用它来表达论证的结构。我们看到，有些论证是有效的，也就是说，它们遵循了正确的逻辑结构；有些论证是可靠的，也就是说，它们的前提在现实世界中确实为真。一个好的论证必须同时是有效且可靠的，也就是从真实的前提出发，按照正确的逻辑步骤，得到同样真实的结论。

这种方法的一个根本想法是，我们可以给每个命题赋予一个“真值”，每个逻辑命题不是真就是真值为假。事实上，在第6讲里我们也看到，计算机在某种意义上也是以类似的方式运作的：它把电信号解释为“开”或“关”，也就是1和0，而它进行的计算，本质上就是由这些输入信号的逻辑组合所构成的运算。简单地说，计算机以逻辑思考，并且以二进制计数。

我们可以把计算机想象成一种“逻辑机器”，它能够接受任意逻辑命题，然后输出True或False的判断。如果我们把 T 换成1，把 F 换成0，那么可以把计算机对命题的判断可视化如下：

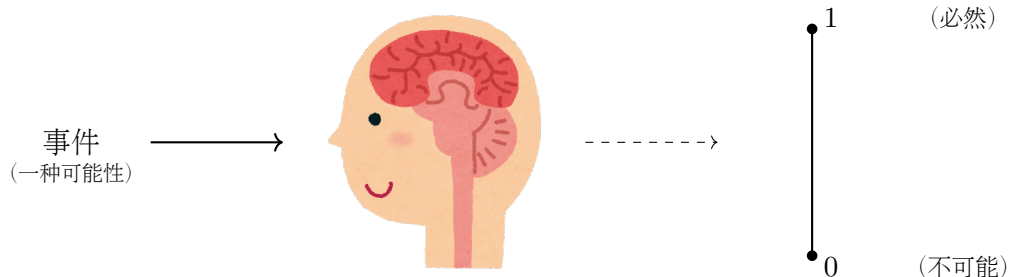


对于涉及数学或电信号这类清晰而精确对象的逻辑命题，计算机非常有用。然而，人类的思维并不完全是这样。现实世界常常迫使我们在并不知道一个命题究竟是真还是假的情况下进行推理。比如，考虑这样一个命题：“明天早上会下雨。”如果我们是气象学家，也许知道答案；但我们并不是气象学家，所以我们能诚实说出的最多也只是“我不知道，也许吧。”从这个意义上说，问题不在于这个命题没有真值，而在于我们不知道哪个真值才是对的。换句话说，世界对于我们来说可以是不确定的，但我们依然能够在这种不确定性之下进行推理。

1.2 概率论

概率论是不确定性的数学研究。它与逻辑不同，区别在于：逻辑告诉我们，在给定前提的情况下，哪些结论必然成立；而概率则在信息不完整时，为我们提供一种衡量不同结果有多大可能发生的方法。从某种意义上说，概率论可以看作是把“相信的程度”进行数量化的科学。

在本课程中，我们把一次试验（例如掷骰子或抛硬币）看作有若干个可能结果的过程，但在试验发生之前，我们并不知道究竟会出现哪一个结果。概率会给这些可能结果中的事件分配一个介于0和1之间的数，我们把这个数理解为该结果有多“可信”或“可能”。概率论在这个基础上进一步发展出一套巧妙的规则，使我们能够根据已有信息，或者根据试验本身的设定，来确定这些数值。和前面提到的计算机不同，我们这种带有不确定性的头脑，更像是把不可预测的结果看成一条连续的刻度：



需要注意的是，概率**不是**介于0 和1 之间的“真值光谱”。¹ 概率真正表示的是确定程度的光谱：对于每个事件 A （也就是某些结果的集合），我们可以赋予一个介于0 和1 之间的数 $P(A)$ 。其中， $P(A) = 0$ 表示该事件不可能发生， $P(A) = 1$ 表示该事件一定会发生，而中间的数值则表示我们对于哪个结果会出现所具有的不同程度的不确定性。

1.3 随机性

我们将假设系统中的实际结果是随机的，也就是说，没有一条（对我们而言可用的）规则能决定结果到底会是什么。于是我们可能会问：随机性究竟是什么？真实世界真的随机吗？例如，假设我随机地在1 到10 之间选一个数，比如3。为什么我选了这个数？我真的是随机选的吗？还是说，我其实是由过去所有经验的综合影响，才最终选了3？

换一种说法，假设我抛一枚硬币。结果真的是随机的吗？按理说，它总该在某种程度上由物理定律决定。如果我完全知道自己把硬币弹向空中时所施加的力，能不能确定无疑地算出最后的结果？

关于“真正的随机性”这个问题，背后有很深的哲学意味。有些人相信世界是完全决定论的，也就是说，一切都由过去发生的事和物理定律所决定。另一些人则相信，世界上确实存在真正随机的事物（例如量子力学中的某些观测），但即便在那里，相关问题也依然有争议。无论如何，对我们来说，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。所以，即使现实中的事物并不是真正随机的，它们仍然可能在我们看来像是随机的。正因如此，概率论仍然是有用的，它可以帮助我们在不确定性之中作出决策。

2 概率的基本性质

2.1 把概率看作一个概念

从数学上说，概率是一个函数，它把某个随机过程的可能结果映射到数值上。在概率论中，我们通常考虑结果的集合，这样的集合称为事件，并且一般用大写字母 $A, B, C, \dots$ 来表示。于是， $P(A)$ 就表示事件 A 的概率。一个结果是某一次试验中的某一个具体结果（例如“掷出4”）；所有结果组成的集合叫作样本空间，记作 S ；而事件 A 则是 S 的一个子集，也就是某些结果构成的集合。关于概率，有几条重要的基本规则：

- (1) 对任意事件 A ，总有 $0 \leq P(A) \leq 1$ 。
- (2) $P(A) = 0$ 表示“不可能”， $P(A) = 1$ 表示“必然”。
- (3) 如果 $P(A) < P(B)$ ，那么事件 B 比事件 A 更可能发生。

¹顺便一提，确实存在一种逻辑理论会这样做，它叫作“模糊逻辑”。但这**不是**概率。

在本课程中，我们主要关注古典概率，也就是说，我们有有限个可能结果，并把它们看成等可能的。典型例子包括掷骰子、抛硬币、发扑克牌，或者盲猜一道判断题。

2.2 如何赋予概率

假设某个试验有有限个等可能结果，也就是说，每个结果发生的概率都相同。回忆一下，“事件”只是若干结果所组成的集合。对于任意事件，我们可以用下面的公式来赋予它一个概率：

$$\text{事件的概率} = \frac{\text{有利于该事件的结果个数}}{\text{所有结果的总个数}}.$$

换句话说，我们数出有利结果的个数，再除以所有可能结果的总数。在这个分数里，分子总是不超过分母，因此这个比值一定落在0 到1 之间。

同时要注意，上面的公式与“每个结果等可能”这一假设完全一致。我们可以把单个结果看成大小为1 的集合，于是上面的公式就退化为一个特殊情形：

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

2.2.1 概率的表示方式

概率总是介于0 和1 之间的数，但我们可以用几种等价的方式来表示它。有时，我们选择的写法还会透露关于系统的额外信息。比如，如果事件 A 的概率是50%，那么我们也可以同样地写成：

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

这里， $\frac{1}{2}$ 是既约分数，0.5 是小数，50% 是百分数，而 $\frac{250}{500}$ 是未约分的分数，它们表示的其实是同一个概率。在最后一种写法里，也许我们面对的是一个共有500 种可能性的系统，其中有250 个结果是“有利”的。未约分的分数 $\frac{250}{500}$ 告诉了我们这个信息，而约分后的 $\frac{1}{2}$ 则没有。

2.3 例子：玩具柜

你那位过分热情的表亲有一个柜子，里面放着58 件不同的玩具。他会随机拿出一件给你看。在他拿出来之前，你并不知道他会选哪一件，并且你假设每一件玩具被选中的可能性都相同。



设事件 A 为：“被选中的玩具是一辆蓝色小汽车”。观察上图可知，柜子里有两辆蓝色小汽车。因此：

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

设事件 B 为：“被选中的玩具来自第4排”。从上图中数一数，我们看到第4排共有8件玩具，所以事件 B 包含8个结果。因此：

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

设事件 C 为：“被选中的玩具是绿色的”。在柜子里仔细找过以后，我们看到共有10件绿色玩具。因此：

$$P(C) = \frac{10}{58}.$$

2.4 练习：判断题答案的组合

O'Connell 博士出了一份含有3道判断题的作业。他把所有可能的作答模式全部列出来，然后随机选取其中一种。

样本空间共有8个等可能结果：

$$\{TTT, TTF, TFT, FTT, TFF, FTF, FFT, FFF\}.$$

练习1

回答下列问题。

- (1) 三个答案全为False 的概率是多少？
- (2) 恰好有两个答案为True 的概率是多少？
- (3) 如果一名学生完全盲猜，他把3题全部猜对的概率是多少？

解答

- (1) 只有一个有利结果，即FFF，所以

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

- (2) 有利结果为TTF、TFT 和FTT，共3个，所以

$$P(\text{恰有两个True}) = \frac{3}{8}.$$

- (3) 正确答案模式只有一种，所以

$$P(3 \text{ 题全对}) = \frac{1}{8}.$$

2.5 大数定律

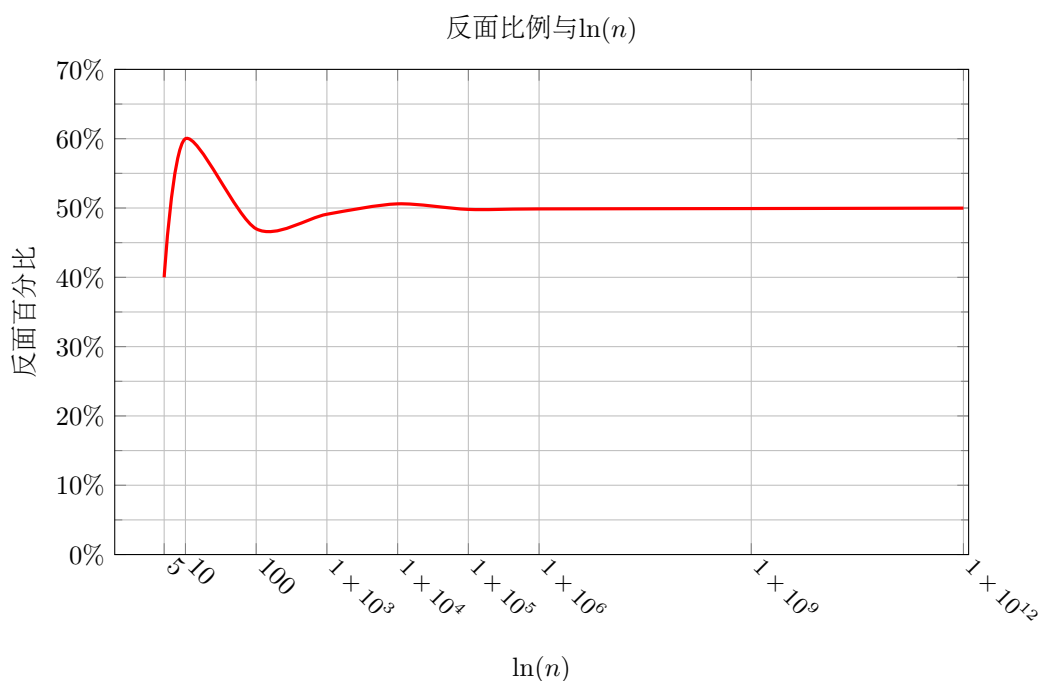
大数定律告诉我们，如果把同一个试验重复很多次，那么所得的平均结果会越来越接近理论上的平均值。举个例子，假设我们有一枚公平硬币。我们预期，抛出反面的概率是50%，也就

是 $P(T) = \frac{1}{2}$ 。按照我们对概率的理解，这意味着从长期来看，我们应当得到大约50% 的反面。

为了看到这一点，假设我们开始一遍又一遍地抛硬币。下面用随机数生成器给出了一个样本中反面所占比例的例子：

- 抛5 次：HHTTH。 $T = 2$ ， $H = 3$ ， 所以反面占40%。
- 抛10 次：HHTTHTHTTT。 $T = 6$ ， $H = 4$ ， 所以反面占60%。
- 抛100 次： $T = 47$ ， $H = 53$ ， 所以反面占47%。
- 抛1000 次： $T = 491$ ， $H = 509$ ， 所以反面占49.1%。
- 抛10000 次： $T = 5062$ ， $H = 4938$ ， 所以反面占50.6%。
- 抛100000 次： $T = 49872$ ， $H = 50128$ ， 所以反面占49.8%。
- 抛100 万次： 反面占49.87%。
- 抛10 亿次： 反面占49.92%。
- 抛1 万亿次： 反面占49.982%。

我们可以把它画成下面的图。这里令 n 表示投掷次数，并使用对数刻度 $\ln(\cdot)$ ，这样不同数量级的投掷次数就可以放在同一条横轴上：



你可以看到，在一开始，结果会有相当大的随机波动。然而随着次数增加，这些随机变化大体上会相互抵消，反面的比例便会逐渐收敛到理论平均值50%。这为概率提供了一种可能的解释：

概率的一种解释

当我们说 $P(A) = x\%$ 时，意思是：如果把这个试验重复极大量次，那么与事件 A 对应的结果所占比例会越来越接近 $x\%$ 。

2.6 重要定义

- (1) 一个简单事件恰好只包含试验的一个结果。
- (2) 样本空间是所有可能结果（简单事件）的集合。
- (3) 一般的事件 A 是样本空间的任意子集，因此它可以包含一个结果，也可以包含多个结果。
- (4) 样本空间中所有简单事件的概率之和总是1。
- (5) 当所有结果等可能时，我们通过数出有利结果的个数，再除以总结果数来计算概率。

为了看到这些定义如何运作，让我们回到例2.3中的玩具柜。在那里，样本空间是

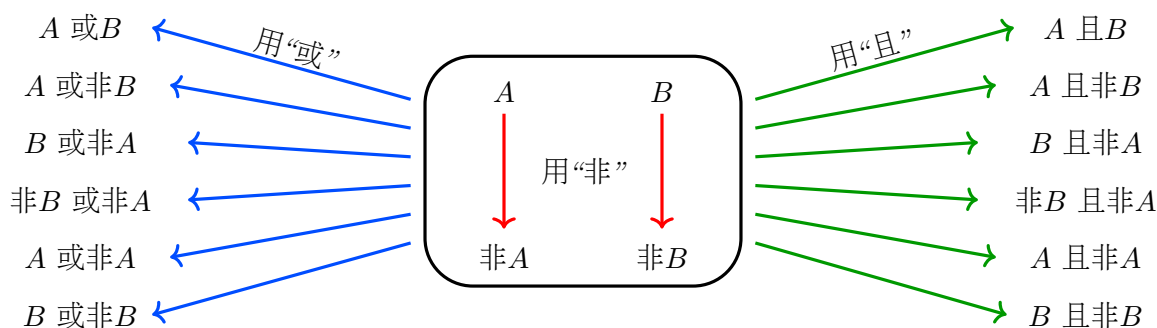
$$S = \{\text{柜子里的所有玩具}\},$$

因此集合 S 的大小是58。在这个例子里，一个结果就是选中某一件具体玩具，也就是“恰好选中这一件玩具”。从数学上说，这就是 S 的一个只含一个元素的子集。一般事件则是 S 的任意子集，它们对应更一般的叙述，例如“选中第四排中的某件玩具”或者“选中一辆蓝色小汽车”。这里要特别强调一点：事件本身就是样本空间的一个子集。比如，在第2.3节里我们考虑了事件 C ，它表示“选中一件绿色玩具”，并且其概率是 $\frac{10}{58}$ 。从数学上看，事件 C 的定义就是柜子里所有绿色玩具所组成的集合。

3 复合概率

3.1 复合事件

假设我们有两个事件 A 和 B 。根据前面的讨论，我们已经知道如何求出 $P(A)$ 和 $P(B)$ ：只需数出与 A 或 B 对应的结果个数，再除以样本空间的大小，也就是所有可能结果的总数。事实上，我们还可以用三个逻辑构件来造出许多别的事件，即“且”“或”“非”：



这些看起来和前几讲里我们见过的命题逻辑联结词非常相似。不过，它们之间仍然有差别。这里我们不是在精确的1和0的世界中工作，而是在研究赋予不确定性的那些数值。因此，虽然上面的图和命题逻辑中的联结词很像，但在概率论里，这些联结词并不完全按照相同的方式运作。事实上，在概率论中，把这些逻辑操作翻译成集合论操作会更加有用。

3.2 集合论运算

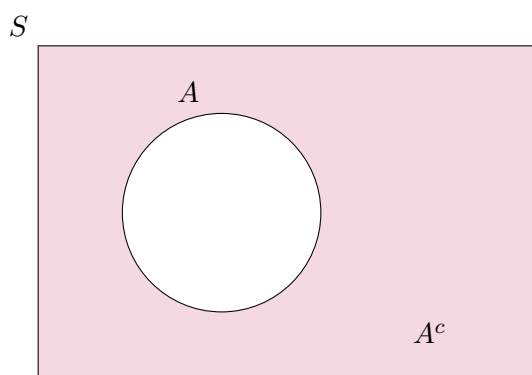
“非”“且”“或”这三个操作可以按如下方式翻译成集合论语言。

- (1) 对任意事件 A ，短语“非 A ”表示 A 不发生的事件。在集合语言中，“非 A ”由 A 的补集给出，记作 A^c 。
- (2) 对两个事件 A 和 B ，“ A 且 B ”表示 A 和 B 同时发生的事件。在集合语言中，“ A 且 B ”由 A 和 B 的交集给出，记作 $A \cap B$ 。
- (3) 对两个事件 A 和 B ，“ A 或 B ”表示 A 发生，或 B 发生，或两者都发生的事件。在集合语言中，“ A 或 B ”由 A 和 B 的并集给出，记作 $A \cup B$ 。

下面我们将依次介绍这些集合运算。

3.2.1 集合的补集

补集 A^c 表示样本空间中所有不属于子集 A 的元素。在下图中，阴影部分就是补集：

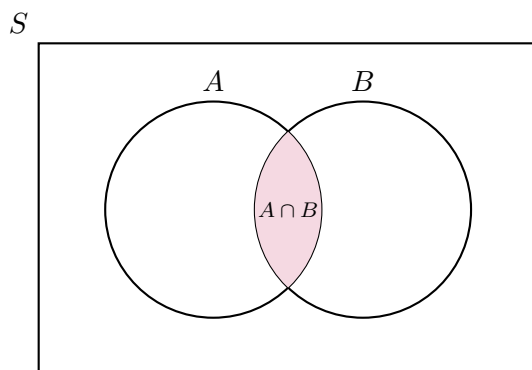


举个例子，设 $S = \{0, 1, 2, 3, \dots\} = \mathbb{N}$ ，也就是自然数集合。如果我们取子集 $A = \{\text{偶数}\}$ ，那么 A 的补集就是 S 中所有不在 A 里的数。换句话说，补集 A^c 由那些在 S 中却缺少于 A 的元素构成。既然 A 是所有偶数的集合，那么补集就是

$$A^c = \{\text{奇数}\}.$$

3.2.2 两个集合的交集

对于两个子集 A 和 B ，它们的交集 $A \cap B$ 是另一个子集，它的元素就是那些同时属于 A 和 B 的对象。“交叉路口”这个词在道路中很常见，它指的是两条路相交的地方。从这个意义上说，如果我们正好站在交叉点上，那么我们就同时站在两条路上。集合论中的交集也是同样的道理：子集 $A \cap B$ 就像两个集合之间的重叠部分，如下图阴影所示：

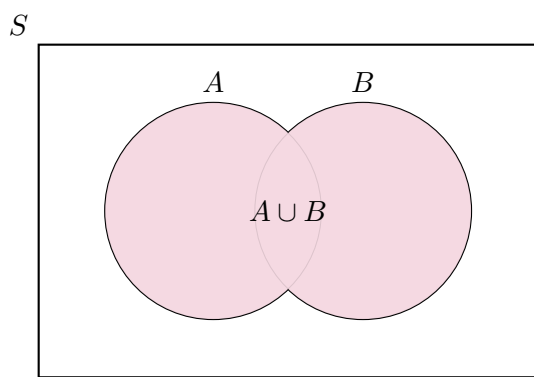


例如，我们再次取 $S = \mathbb{N}$ ，并令 $A = \{\text{偶数}\}$ 。现在再取 S 的另一个子集 $B = \{\text{小于10的数}\}$ 。那么， A 和 B 的交集就是那些同时属于 A 和 B 的数。由于 A 是所有偶数的集合，而 B 是所有小于10的数的集合，所以交集 $A \cap B$ 就是那些同时满足“偶数”且“小于10”的数。写出来就是：

$$A \cap B = \{0, 2, 4, 6, 8\}.$$

3.2.3 两个集合的并集

“并集”这个词通常表示“把东西合在一起”。例如，European Union（欧盟）是由一些欧洲国家联合而成的政治组织。在集合论里，意思类似：对于两个子集 A 和 B ，它们的并集 $A \cup B$ 是一个子集，包含所有属于 A 、或者属于 B 、或者同时属于两者的元素。图像上，可以把它想成把 A 和 B 两个集合全都覆盖起来：



举一个简单例子，设 $A = \{1, 2, 3\}$ ， $B = \{3, 4, 5, 6\}$ 。那么并集 $A \cup B$ 就是把 A 和 B 中所有元素放进同一个大集合后得到的集合：

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

注意，我们不会把3写两次，它只算一次。

3.2.4 练习：集合运算

设 $S = \mathbb{N} = \{0, 1, 2, 3, 4, \dots\}$ ，即自然数集合。考虑两个子集：

$$A = \{\text{偶数}\}, \quad B = \{\text{小于10的数}\}.$$

练习2

求下列各项。

- (1) B^c (B 的补集)。
- (2) $A \cap B$ 。
- (3) $A^c \cap B$ 。

解答

(1)

$$B^c = \{\text{大于或等于10 的数}\}.$$

(2)

$$A \cap B = \{0, 2, 4, 6, 8\}.$$

(3) A^c 是奇数集合, 所以

$$A^c \cap B = \{1, 3, 5, 7, 9\}.$$

3.3 概率 $P(A^c)$

对于任意事件 A , 事件 A 和事件 A^c 一起覆盖整个样本空间, 并且彼此没有重叠。用集合语言来说, $A \cup A^c = S$, 且 $A \cap A^c = \emptyset$ 。所以它们的概率之和等于1:

$$P(A) + P(A^c) = 1.$$

因此:

$$P(A^c) = 1 - P(A).$$

3.3.1 练习: FFF 的补事件

在前面的判断题试验中, $P(\text{FFF}) = \frac{1}{8}$ 。

设 E 为事件“出现FFF”。那么 E^c 就表示“不是FFF”, 也就是其余7 个结果中的任意一个。

练习3

求 $P(E^c)$ 。

解答

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

3.4 概率 $P(A \cap B)$

在我们的“古典概率”设定中, 概率 $P(A \cap B)$, 也就是事件“ A 且 B ”的概率, 应当通过数交集的大小再除以样本空间的大小来确定, 即

$$P(A \cap B) = \frac{A \cap B \text{ 的大小}}{S \text{ 的大小}}.$$

在大多数简单情形下, 我们可以直接靠计数来计算 $P(A \cap B)$ 。例如, 假设我们掷一枚公平骰子, 并令事件 A 为“掷出偶数”, 令事件 B 为“掷出的数小于4”。那么, 计算 $P(A \cap B)$ 时, 我们只需把有利结果明确列出来即可。在1 到6 之间, 同时满足“小于4”且“是偶数”的只有一个数, 那就是2。因此, 子集 $A \cap B = \{2\}$ 的大小为1, 相应的概率就是

$$P(A \cap B) = \frac{1}{6}.$$

上面的例子之所以容易，是因为样本空间只有6个元素。然而在更复杂的情形下，要把所有可能情况都列出来，并准确数出交集 $A \cap B$ ，会变得非常困难。实际中，更有帮助的方法通常是使用一些公式，把 $P(A \cap B)$ 与 $P(A)$ 和 $P(B)$ 联系起来。假设我们知道 A 和 B 的大小，那么就总能用前面类似的计数公式算出 $P(A)$ 和 $P(B)$ 。但是，仅仅知道 A 和 B 的大小，并不能决定它们交集的大小。因此，计算 $P(A \cap B)$ 会稍微细致一些，而且很大程度上取决于这两个事件彼此之间的关系。

3.4.1 条件概率

要想求出交集 $A \cap B$ 的大小，一种办法是“进入”集合 B 的内部，然后数一数其中哪些元素也属于 A 。用概率语言来说，这就相当于在样本空间从整个 S 缩小为子集 B 的前提下，去考虑事件 A 发生的概率。换言之，你是在评估：已知 B 会发生时， A 发生的可能性有多大。这个想法就叫作条件概率，记作 $P(A|B)$ ，读作“在 B 已发生的条件下， A 的概率”。

回到掷骰子的例子：事件 A 表示“掷出偶数”，事件 B 表示“掷出的数小于4”。想象一下，在你掷骰子之前，上帝在你耳边悄悄说了一句：“结果会小于4。”这意味着，在那一刻，你会意识到不可能掷出4, 5, 6。在这种情况下，如果你要重新评估事件 A 发生的机会，你就会把可能性空间从 $\{1, 2, 3, 4, 5, 6\}$ 缩小到更小的集合 $\{1, 2, 3\}$ ，而这个集合正好就是 B 。因此，当你重新评估可能性时，所做的计数是

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

与 A 和 B 对应的子集分别是 $\{2, 4, 6\}$ 和 $\{1, 2, 3\}$ 。因此，交集 $A \cap B = \{2\}$ 的大小为1。由于 B 的大小为3，所以条件概率变为

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

现在我们做一个有用的观察。在这个例子中，我们已经看到

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

正好有

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

用符号写出来，我们刚刚发现了一个一般公式的具体例子：

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

因此，要计算复合事件“ A 且 B ”的概率，我们总可以利用一个包含条件概率的乘法公式。等价地，也可以写成

$$P(A \cap B) = P(A) \cdot P(B | A).$$

3.4.2 概率独立性

在某些情况下， $P(A \cap B)$ 的公式可以大幅简化。为了准确说明这一点，我们需要区分独立事件和相关事件。

独立事件与相关事件

设 A 和 B 是两个事件。我们说：

- 如果知道 A 发生并不会给你任何关于 B 是否发生的信息，那么 A 和 B 是**独立的**；
- 如果知道 A 发生会给你一些关于 B 是否发生的信息，那么 A 和 B 是**相关的**。

提醒一下：概率研究的是不确定性。因此，相关事件背后的思想是，知道事件 A 发生，可能会改变事件 B 的概率。只要关于 A 的信息与 B 有某种关联，这种情况就可能发生。

在前面的掷骰子例子里，事件 A 为“掷出偶数”，事件 B 为“掷出的数小于4”，它们是相关事件，因为知道其中一个事件发生，会影响你对另一个事件可能性的判断。不过，我们也能想象别的情形，在那里知道一条额外信息，并不会给我们任何理由去改变原先的概率判断。例如，令事件 C 为“掷出的数是3 的倍数”。则作为子集，

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

如果这时上帝在你耳边说：“结果会是3 的倍数。”那么你就会把样本空间从 $\{1, 2, 3, 4, 5, 6\}$ 缩小到 $\{3, 6\}$ 。在这个新的受限集合中，一个是奇数，一个是偶数，因此掷出偶数的概率依然是 $\frac{1}{2}$ 。这与 $P(A)$ 相同，所以知道 C 会发生，并没有改变你对系统所具有的确信程度。因此，事件 A 和 C 是独立的。

按照这种思路，如果额外信息并不会改变你原本的不确定性，那么这两个事件就是独立的。换句话说，当条件概率与无条件概率相同时，两个事件独立。也就是说，若

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

则事件 A 和 B 独立。这会把 $P(A \cap B)$ 的公式化简为下式。

独立事件的乘法法则

如果 A 和 B 是独立事件，那么：

$$P(A \cap B) = P(A) \cdot P(B).$$

例：考虑一枚公平硬币，被抛两次。令事件 A 为“第一次抛出反面”，令事件 B 为“第二次抛出反面”。这两个事件是独立的，因为第一次抛出的结果并不会给你关于第二次抛掷结果的任何信息。因此，我们可以使用独立事件的公式，得到：

$$P(A \cap B) = \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) = \frac{1}{4}.$$

等价地说，两次抛掷的样本空间是

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

而只有 TT 是有利结果，所以概率是 $\frac{1}{4}$ ，与上面的计算一致。

3.4.3 例：独立事件（两副牌）

假设我取两副彼此分开的扑克牌，并分别把它们洗匀。令

$$A = \text{“第1 副牌的顶牌是A”}, \quad B = \text{“第2 副牌的顶牌是A”}.$$

这两个事件是独立的，而且每个事件的概率都是

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

因此

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

3.4.4 例：相关事件（一副牌中连续两张A）

假设我把一副52 张的扑克牌洗匀。令

$$A = \text{“第一张牌是A”}, \quad B = \text{“第二张牌是A”}.$$

这两个事件是相关的，因为如果第一张牌已经是A，那么牌堆中剩下的A 就更少了。

计算得：

$$P(A) = \frac{4}{52} = \frac{1}{13}, \quad P(B | A) = \frac{3}{51}.$$

因此：

$$P(A \cap B) = \left(\frac{1}{13}\right) \left(\frac{3}{51}\right) = \frac{1}{221}.$$

3.4.5 乘法法则总结

步骤1: 判断事件是否独立。

（假设 B 不是不可能事件。）如果 $P(A) = P(A | B)$ ，那么它们独立。

步骤2: 使用正确的乘法法则。

如果独立，则 $P(A \cap B) = P(A) \cdot P(B)$ 。

如果不独立，则 $P(A \cap B) = P(A) \cdot P(B | A)$ 。

3.5 概率 $P(A \cup B)$

根据第3.2 节，复合事件“ A 或 B ”对应于集合论中的并集 $A \cup B$ 。因此，要计算概率 $P(A \cup B)$ ，我们可以数出这个子集的大小，再除以样本空间的大小：

$$P(A \cup B) = \frac{A \cup B \text{ 的大小}}{S \text{ 的大小}}.$$

就像前面讨论交集时那样，一般来说，我们不能只凭 A 和 B 的大小，就推出并集 $A \cup B$ 的大小。我们还需要知道它们彼此之间如何重叠。幸运的是，我们刚刚已经得到了一个计算重叠部分的方法，因此实际上可以把并集 $A \cup B$ 的大小与 A 、 B 以及交集 $A \cap B$ 的大小联系起来。公式是：

$$A \cup B \text{ 的大小} = (A \text{ 的大小}) + (B \text{ 的大小}) - (A \cap B \text{ 的大小}).$$

这个公式之所以成立，是因为我们想要把 A 和 B 中的所有元素都算进去。于是，直接把 A 的大小和 B 的大小相加，看起来似乎很合理。但如果这么做，就有可能把交集 $A \cap B$ 中的元素重复计算。你可以亲自看看这个例子：设

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

这两个集合的大小都为3，所以大小之和是 $3 + 3 = 6$ 。然而，并集

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

的大小其实是4，而不是6。当我们计算 $3 + 3 = 6$ 时，元素2和3实际上被重复算了两次。但如果把多算的那一份 $\{2, 3\}$ 去掉，就得到 $6 - 2 = 4$ ，这才是并集的正确大小。

利用这个观察，我们可以推出一个关于 $P(A \cup B)$ 的一般公式，用 $P(A)$ 、 $P(B)$ 和 $P(A \cap B)$ 来表示：

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

有时候，事件 A 和 B 所对应的子集可能完全没有重叠。在这种情况下，交集 $A \cap B$ 是空集（也就是大小为零），因此就不存在任何重复计算的风险。每当这种情况发生时，我们就称事件 A 和 B 是互斥的。由于 A 和 B 的子集对应于使事件 A 和 B 实现的结果集合，互斥事件就不可能同时发生，也就是说

$$P(A \cap B) = 0.$$

如果 A 和 B 是互斥事件，那么 $P(A \cup B)$ 的公式就简化为

$$P(A \cup B) = P(A) + P(B).$$

3.5.1 例：A、黑桃、K（顶牌）

假设我把一副扑克牌洗匀。令

$$A = \text{“顶牌是一张A”}, \quad B = \text{“顶牌是一张黑桃”}, \quad C = \text{“顶牌是一张K”}.$$

A 和 B 不是互斥的，因为黑桃A同时属于 A 和 B 。 A 和 C 是互斥的，因为顶牌不可能同时既是A又是K。

练习4

求下列各项。

- (1) $P(A \cup B)$ 。
- (2) $P(A \cup C)$ 。

解答

(1)

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

所以

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

(2) A 和 C 互斥, 所以

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

4 练习

4.1 掷骰子

假设我们掷一枚公平的六面骰子。令

$$A = \text{“掷出偶数”}, \quad B = \text{“掷出的数} \geq 4\text{”}.$$

练习5

求下列各项。

(1) $P(A)$ 。

(2) $P(B)$ 。

(3) $P(A \cap B)$ 。

(4) $P(A \cup B)$ 。

解答

(1) $A = \{2, 4, 6\}$, 所以

$$P(A) = \frac{3}{6} = \frac{1}{2}.$$

(2) $B = \{4, 5, 6\}$, 所以

$$P(B) = \frac{3}{6} = \frac{1}{2}.$$

(3) $A \cap B$ 表示“偶数且 ≥ 4 ”, 所以为 $\{4, 6\}$, 因此

$$P(A \cap B) = \frac{2}{6} = \frac{1}{3}.$$

(4) $A \cup B$ 表示 $\{2, 4, 5, 6\}$, 所以

$$P(A \cup B) = \frac{4}{6} = \frac{2}{3}.$$