

作者	生姜 DrGinger
脚本	生姜 DrGinger
视频	崔崔 CuiCui
开源学习资源	https://github.com/Visualize-ML
平台	https://www.youtube.com/@DrGinger_Jiang https://space.bilibili.com/3546865719052873 https://space.bilibili.com/513194466

1.3 概率



本节你将掌握的核心技能：

- ▶ 概率是衡量事件发生可能性的数值，范围在 0 到 1 之间。
- ▶ 在古典模型中，每个样本点发生的可能性相同。
- ▶ 用枚举法计算事件发生的概率。
- ▶ 试验次数增加时，频率趋近理论概率。
- ▶ 掌握和事件、互斥事件与对立事件的概念及概率计算方法。

本节介绍概率的基本概念及其计算方法。我们将学会如何理解概率的含义，掌握古典概率模型中“等可能”原则，通过枚举法计算事件概率。我们还区分了理论概率与试验概率，并通过模拟演示了大量试验中概率趋于稳定的规律。

概率

简单来说，**概率** (probability) 是对不确定性事件发生可能性的度量。

概率把原本模糊的“不确定”转化为一个介于 0 和 1 之间的数值。

如图 1 所示，概率为 1 意味着事件必然发生，概率为 0 意味着事件不可能发生。概率越接近 1，表示该事件发生的可能性越大；越接近 0，表示该事件发生的可能性越小。

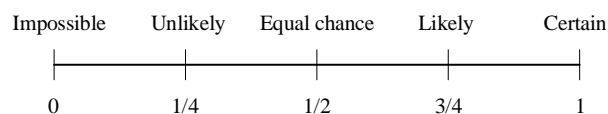


图 1.0 ~ 1 概率取值

等概率

在很多体育运动中，比如网球、羽毛球各类比赛开始前，常常会用抛硬币来决定谁先发球或者挑边。硬币落地后是正面还是反面，每个结果出现的机会看起来都是一样的，因此所有运动员都心服口服，没人会质疑这种方法不公平，也不会说有什么猫腻。

大家之所以接受这种方式，是因为在我们的基本认知里，抛硬币正反面是等可能的，也就是说正面和反面出现的概率都是 0.5。

这种基于对称性和经验的判断，是最直观的等概率思想。正是因为有了“等可能”，我们才能很自然地算出抛硬币正面朝上的概率是 $1/2$ 。图 2 所示为用火柴梗图可视化这个随机试验的概率值。火柴梗图每个线段的水平位置代表可能的取值，线段的高度代表该取值对应的概率值。

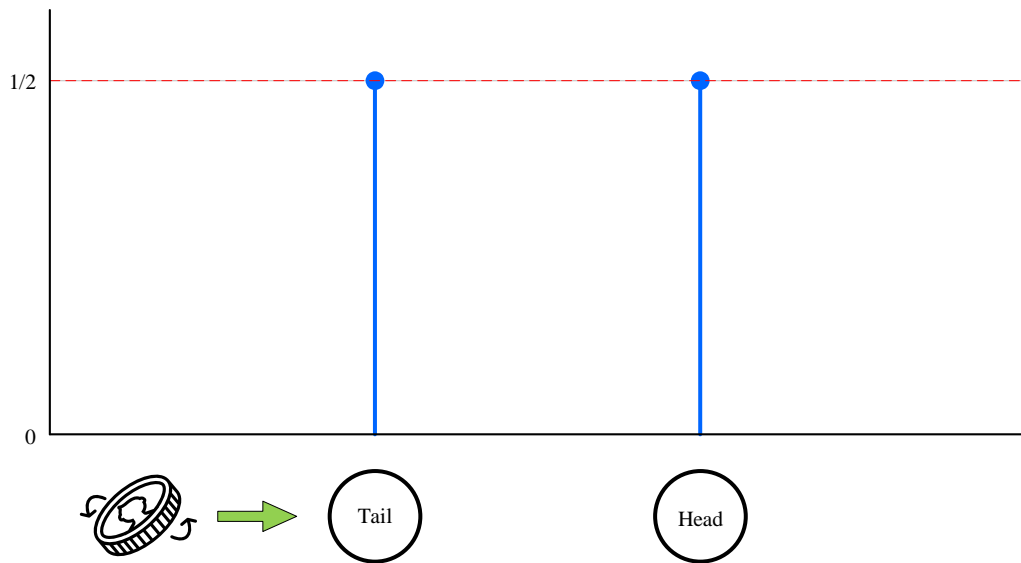


图 2. 概率值火柴梗图，抛一枚质地均匀硬币

把这种“等概率”的推理方式推广到更多随机试验，就是经典概率论的出发点。

设随机试验的样本空间 Ω 是有限集合，事件 $A \subset \Omega$ 。如果 Ω 的每个样本点发生的可能性相同，则称

$$\Pr(A) = \frac{\text{card}(A)}{\text{card}(\Omega)} \quad (1)$$

为事件 A 的概率。 $\text{card}(A)$ 、 $\text{card}(\Omega)$ 分别表示事件 A 和样本空间 Ω 中样本点的个数。

能够用 (1) 描述的概率模型称为**古典概率模型** (Classical Probability Model)，简称为古典概型。

简单来说，古典概型，是指在样本空间中，每个样本点发生的可能性都相等的概率模型。

举一个熟悉的例子，掷一颗均匀色子，它的样本空间是 $\Omega = \{1, 2, 3, 4, 5, 6\}$ ，其中每一个点数出现的可能性都是相同的，都是 $1/6$ 。

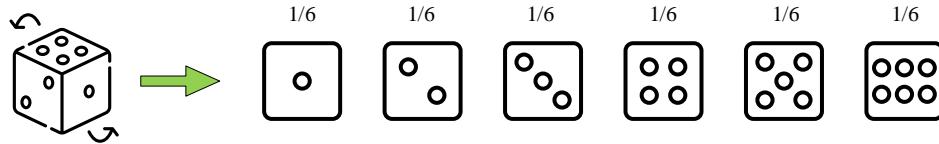
图 3. 掷一颗色子，每个点数出现的概率都是 $1/6$

图 3 的概率值也可以用如图 4 所示的火柴梗图可视化。

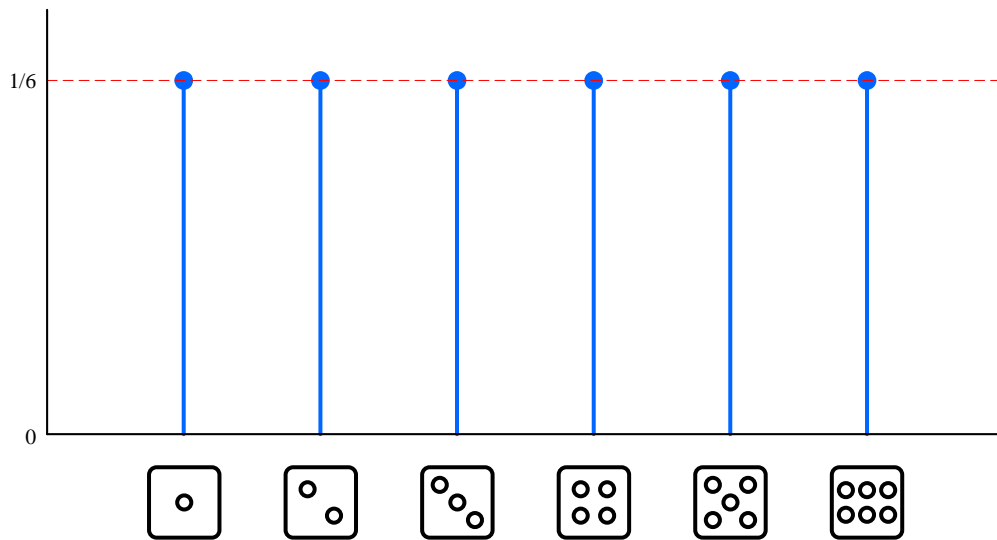


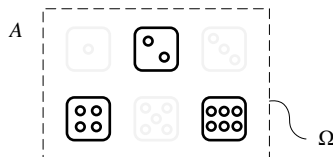
图 4. 概率值火柴梗图，掷一颗质地均匀色子

如图 5 所示，掷一颗均匀色子，定义事件 A 为“点数是偶数”的概率， $A = \{2, 4, 6\}$ ，有 3 个可能结果，因此

$$\Pr(A) = \frac{\text{card}(A)}{\text{card}(\Omega)} = \frac{3}{6} = \frac{1}{2} \quad (2)$$

相信大家已经看到，在古典概型中，概率的计算就变成了“目标情况数除以总情况数”，即枚举法。

? 掷一颗均匀色子，事件 A 为点数大于 3，请计算事件 A 的概率。

图 5. 掷一颗均匀色子，事件 A 为结果点数为偶数

古典概率模型的核心思想在于“等可能性假设”。它适用于理想化的、对称性明显的实验，例如掷硬币、掷骰子、从一副洗匀的扑克牌中抽牌等。在这些情况下，我们能够合理地假设每个结果出现的机会相同。

总结来说，概率是用来衡量事件发生可能性大小的数值，因此它必须满足一些基本的数学性质。首先，概率总是非负的，即对于任意事件 A ，都有

$$\Pr(A) \geq 0 \quad (3)$$

这意味着事件的发生不可能具有“负的可能性”。

其次，**必然事件** (certain event) 的概率为 1。必然事件是指在随机试验中一定会发生的事件，它对应的集合就是整个样本空间

$$\Pr(\Omega) = 1 \quad (4)$$

比如，掷一颗色子，定义事件 A 为结果点数大于 0，显然 $A = \{1, 2, 3, 4, 5, 6\}$ ，即整个样本空间 Ω 。无论掷多少次，结果一定落在这个集合中，因此 $\Pr(A) = 1$ 。

相对地，**不可能事件** (impossible event) 的概率为 0。不可能事件是指无论试验如何进行都不会发生的事件，它对应的是**空集** (empty set)

$$\Pr(\emptyset) = 0 \quad (5)$$

再如，同样掷一颗色子，定义事件 A 为结果点数大于 6，显然 $A = \{\}$ ，即 A 为空集。因此， $\Pr(A) = 0$ 。

! 再次强调，概率是对“发生可能性”的度量，其本质决定了它必须非负。

理论概率、试验概率

本节前文给出的是用古典概率模型（等可能事件和枚举法）得出的**理论概率** (theoretical probability)。也称为公式概率或数学概率，是一种基于理论推导的概率计算方法。

根据大量的、重复的统计试验结果计算随机事件中各种可能发生结果的概率，称为**试验概率** (experimental probability)。

理论概率可以作为试验概率的基础，即在假设所有可能的结果是等可能的情况下，理论概率可以预测事件发生的概率，而试验概率则可以验证这一预测是否准确。

例如，大量抛一枚“质地均匀”的硬币，正面朝上的次数会接近总次数的一半；也就是说，如得到正面的概率不断接近 $1/2$ 。

想象一下，我们手里有一枚均匀的硬币（正反面概率各为 50%）。只抛 10 次：我们可能抛出了 7 次正面，3 次反面。正面的频率是 70%，和理论上的 50% 差得很远。这很正常，因为次数太少，偶然性占了上风。抛 100 次：正面次数可能变成了 55 次，反面 45 次。正面频率降到了 55%，开始向 50% 靠拢。

我们做了一个简单的模拟，结果如图 6 所示。连续抛掷一枚硬币 500 次。每次抛掷的结果不是正面（红点 ●，用数字 1 表示），就是反面（蓝点 ●，用数字 0 表示），且每个结果出现的概率理论上都是 50%。随着抛硬币次数不断增多，我们发现正面出现了约一半的次数，反面也差不多一样多，虽然具体数字会因为随机性有一些波动。

这种模拟可以帮助我们直观地理解“等概率事件”与“试验概率趋近于理论概率”的概念——当抛掷次数足够多时，正反两面的出现频率会越来越接近 $1/2$ ，也就是 50%。这正是概率在实际中的一个体现。例如，如果我们实际掷骰子 600 次，其中有 302 次出现偶数点，那么事件“掷出偶数”的试验概率约为 0.503。

⚠ 注意，图 6 中模拟的随机试验还是——抛一枚硬币。这个随机试验的样本空间依然是 {正面，反面}。只不过我们将这个随机试验反复进行 500 次。

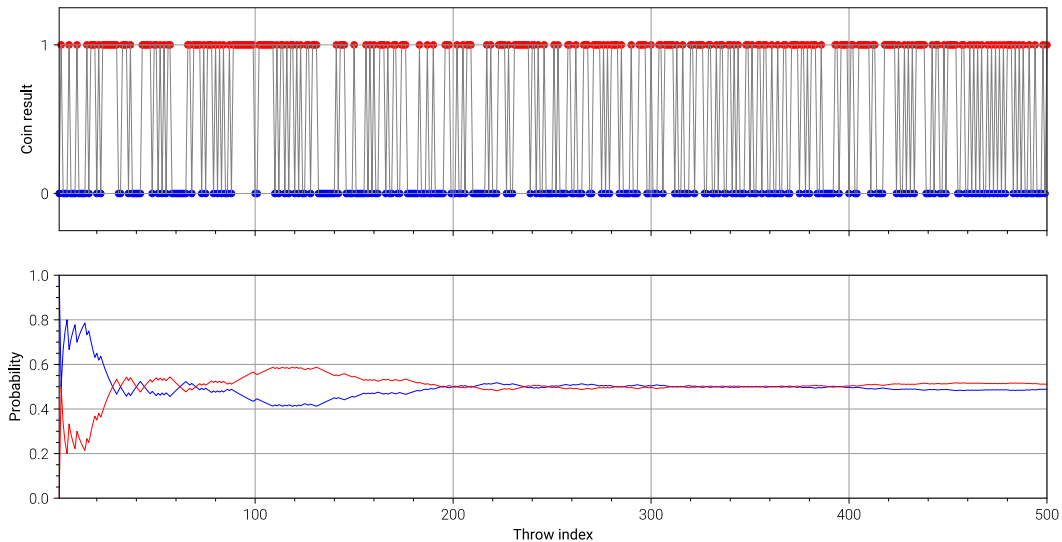


图 6. 抛一枚硬币 500 次，模拟结果

尽管单次随机事件的结果不可预测，但在大量重复试验中，随机事件会呈现出一定的统计规律。随机正是“必然中包含偶然，偶然中蕴含必然”的体现，是自然界和日常生活中普遍存在的一种现象。

从长期来看，当实验次数足够多时，试验概率会逐渐趋近于理论概率，这一现象被称为**大数定律** (Law of Large Numbers)。它揭示了一个重要事实：虽然单次随机事件是不可预测的，但通过重复实验，我们能够发现随机性背后的统计规律。从生活的角度来看，大数定律告诉我们，个例的成败不足喜，也不足悲；只有经过足够多的重复，事物的本质才会真正显现出来。

试想，抛均匀硬币试验中，前面连续 5 次都是正面，是否第 6 次抛掷得到反面的概率会大大升高，大于 $1/2$ ？

实际上，在一次独立重复的抛掷均匀硬币试验中，每一次结果都与之前的结果相互独立，这一点在数学上意味着各次试验之间不存在条件依赖关系。

即使前面连续 5 次都出现正面，这一历史结果不会改变第 6 次抛掷的概率结构。对于一枚均匀硬币而言，每一次抛掷得到正面或反面的概率始终都是相同的，均为 $1/2$ 。因此，第 6 次出现反面的概率仍然是 $1/2$ ，而不会因为前面连续出现正面而“补偿性”升高。

认为连续出现正面后反面“更有可能出现”的直觉是一种典型的认知偏误，称为“赌徒谬误”，其本质是错误地将独立事件当作具有某种短期平衡机制的过程。事实上，只有在大量重复试验的整体层面上，正反面出现的频率才会趋于各占一半，而这种长期规律并不会对单次试验的概率产生任何影响。

换一种更直白的说法：很多人会觉得“前面已经连着出了 5 次正面，接下来该轮到反面了吧”，但这其实是一种错觉。硬币没有记忆，它不会“记得”前面发生了什么，本质上就不是“守得云开见月明，静待花开终有时”那一类过程。第 6 次抛的时候，就像是一次全新的开始，正反面的概率还是各占一半。

从长期来看，如果你抛很多很多次，比如成千上万次，那么正面和反面的比例会慢慢接近一半一半。这时候，前面那“连续 5 次正面”的影响，会被后面海量的随机结果“淹没掉”。也就是说，大量的数据会把这种短期的“偏差”平均掉，让整体看起来是均衡的。

但关键在于，这种“均衡”是靠长期累积实现的，而不是靠下一次抛掷去“纠正”。第 6 次并不会因为前面偏了，就更倾向于出反面；它依然是一个完全独立的随机结果，概率还是 $1/2$ 。

掷色子

假设我们掷一颗“质地均匀”的六面骰子，从理论上讲，由于每个面出现的可能性相同，按照古典概率模型，每个点数出现的概率应为 $1/6$ 。

然而，理论模型只是理想化的描述。为了验证这一结论，我们可以通过计算机模拟或实际操作进行多次试验。例如，连续掷一颗六面骰子 500 次，并记录每一次的结果。每次掷出的点数可能是 1、2、3、4、5、6 之一，且每个结果的出现完全由随机性决定。

如图 7 所示，当我们统计 500 次掷骰结果时，可以发现：虽然六个点数出现的频率并不完全相同——某些点可能稍微多出现几次，而另一些稍少——但整体分布依然非常接近理论预测的 $1/6$ 。这种微小的偏差并不是模型错误，而是随机波动造成的自然现象。如果继续增加实验次数，例如掷 5,000 次或 50,000 次，这些频率之间的差异会逐渐缩小，所有点数的相对频率将越来越接近 $1/6$ 。

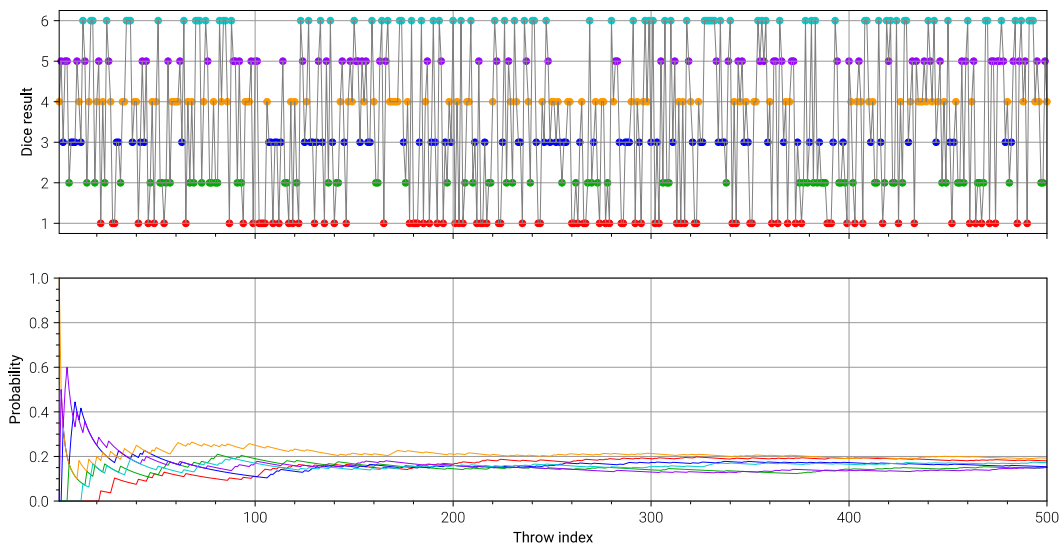


图 7. 掷一颗色子 500 次，模拟结果

代码 1 完成图 6 随机试验模拟，并可视化结果。下面聊聊其中关键语句。

a 引入了两个常用的 Python 模块。matplotlib.pyplot 是用来画图的工具包，简称 plt；numpy 是做数学运算（特别是线性代数运算）的库，比如生成随机数、数组处理等，简称 np。之后我们会频繁调用。

b 设置了随机种子，意思是：每次运行程序时，随机数的结果都一致。这样我们能确保“抛色子”的结果是可重复的。如果不设置种子，每次运行的随机结果可能都不同。建议大家用 # 加在这句代码前“注释掉”，多运行几次代码，观察结果变化。

▲ 在 JupyterLab 中，“注释”、“取消注释”的快捷键为 `ctrl /`。

c 首先定义变量 `n_throws`，就是我们要进行试验的次数。然后用 `np.random.randint(1, 7, size=500)` 模拟 500 次掷色子。`np.random.randint(1, 7, size=500)` 表示从 1 到 6 之间（不包括 7）的整数中，随机取 500 个数，组成一个数组。因为一个正常的六面色子只有 1 到 6 这六个面，所以我们用这个函数模拟掷出的点数。

d 用 `np.arange(1, 501)` 生成从 1 到 500 的一个数组，用来表示每一次抛掷的编号，即第几次。

e 创建了一个字典，叫做 `colors_map`，用来把不同点数映射成不同的颜色。比如，点数是 1 就显示红色，点数是 2 就显示绿色，以此类推。这是为了后面画图时能用不同颜色来区分每个点数。

f 这行代码使用列表生成式 (list comprehension)，把之前掷出的 500 个数逐个映射为颜色，结果是一个包含 500 个颜色值的列表。比如，如果第 1 次是 3，那就变成蓝色；第 2 次是 6，那就变成青色。这个颜色列表后面会用于散点图上给点上色。

在 Python 中，列表生成式是一种简洁的语法形式，用于快速生成新的列表。它的语法形式为 `[expression for item in iterable if condition]`，其中 `expression` 表示要生成的元素，`item` 表示迭代的变量，`iterable` 表示迭代的对象，`if condition` 表示可选的过滤条件。

g 用 `np.zeros()` 创建了一个全是 0 的二维数组，大小是 500 行 6 列。每一行表示前 `n` 次抛色子的统计情况，6 列分别表示点数 1 到 6 出现的频率。随着抛掷次数增加，我们会逐行更新这个表格。

h 是一个 for 循环，从第 1 次到第 500 次，每次都统计目前为止各点数出现的次数。


i 中，`throws[:i+1]` 是前 `i` 次的结果；`np.bincount()` 统计每个点数出现了几次；`minlength=7` 是因为色子的点数范围是 1 到 6，我们加上 1 是为了让点数 6 也能统计进去。`[1:]` 的意思是跳过 0 这个没用的统计项。

j 最后把每个点数的出现次数除以当前的总次数，得到频率，存在第 `i` 行里。

k 为可视化部分。其中，`plt.subplots()` 创建了一个图像窗口，包含两个上下排列的子图 `ax1` 和 `ax2`。`figsize=(12, 6)` 表示整张图的大小，单位是英寸。`height_ratios=[1, 1]` 表示上下图各占一半高度。`sharex=True` 表示上下图的横轴是共用的。

l 绘制图 7 上图。首先用 `plot()` 在 `ax1` 画出一个灰色的细折线，连接每一次掷色子的结果。目的是让我们看到点数的波动趋势。然后用 `scatter()` 在同一幅图上画出彩色的点，表示每次掷色子的具体结果。用的是我们前面生成的 `colors` 颜色列表，每个点颜色和点数一一对应，`s=20` 表示点的大小。剩下的语句为图像装饰。

m 绘制图 7 下图。下图 `ax2` 中先用 for 循环画出六条曲线。每一条表示某个点数的频率变化情况。横轴是抛掷次数，纵轴是当前频率。`cumulative_probs[:, i]` 表示第 `i` 列，也就是点数为 `i+1` 的频率变化过程。

代码 1. 模拟掷一颗色子 |  SP_Ch01_03_01.ipynb

```

## 初始化
a import matplotlib.pyplot as plt
import numpy as np

## 模拟掷色子
# 设置随机种子
b np.random.seed(88)

# 模拟抛掷500次
c n_throws = 500
throws = np.random.randint(1, 7, size=n_throws)

# 抛掷编号
d x_vals = np.arange(1, n_throws + 1)

# 点数颜色映射
e colors_map = {1: '#FF0000', 2: '#00AA00', 3: '#0000FF',
f               4: '#FF9900', 5: '#9900FF', 6: '#00CCCC'}
colors = [colors_map[val] for val in throws]

## 计算累积概率
g cumulative_probs = np.zeros((n_throws, 6))

h for i in range(n_throws):
i     counts = np.bincount(throws[:i+1], minlength=7)[1:] # 忽略0
j     cumulative_probs[i] = counts / (i + 1)

## 可视化
k fig, (ax1, ax2) = plt.subplots(2, 1, figsize=(12, 6),
                                height_ratios=[1, 1], sharex=True)

# 上图
l ax1.plot(x_vals, throws, color='gray', linewidth=0.5, alpha=0.3)
ax1.scatter(x_vals, throws, c=colors, s=20)
ax1.set_ylabel('Dice result')
ax1.set_yticks([1, 2, 3, 4, 5, 6]); ax1.set_xlim(1, x_vals.max())
ax1.grid(True)

# 下图
m for i in range(6):
    ax2.plot(x_vals, cumulative_probs[:, i],
             color=colors_map[i+1])
ax2.set_xlabel('Throw index'); ax2.set_ylabel('Probability')
ax2.set_xlim(1, x_vals.max()); ax2.set_ylim(0, 1); ax2.grid(True)

```

上一节我们已经学习了多个事件之间的关系，例如和事件、积事件、联合概率等。这些概念为我们提供了理解事件间逻辑关系的基础。接下来，我们将把它们与古典概型结合起来，利用集合运算计算不同事件的概率。

和事件、积事件

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

本书配套微视频均发布在 B 站——生姜 DrGinger：<https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：jiang.visualize.ml@gmail.com

定义 A 与 B 为样本空间 Ω 中的两个事件。

上一节提过，事件 $A \cup B$ 为 A 和 B 的和事件。具体来说，当事件 A 和事件 B 至少有一个发生时，事件 $A \cup B$ 发生。 $\Pr(A \cup B)$ 代表事件 A 和 B 和事件概率为

$$\underbrace{\Pr(A \cup B)}_{\text{Union}} = \Pr(A) + \Pr(B) - \underbrace{\Pr(A \cap B)}_{\text{Joint}} \quad (6)$$

其中， $\Pr(A \cap B)$ 为 A 和 B 积事件概率，代表事件 A 和事件 B 同时发生。

上式告诉我们，如果直接将 A 和 B 的概率相加，会把“同时属于 A 和 B 的部分”重复计算一次。因此，(6) 中减去了 $\Pr(A \cap B)$ 。

本书中， $\Pr(A \cap B)$ 常记作 $\Pr(A, B)$ ，常叫做**联合概率** (joint probability)，表示两件事情同时满足的可能性。

举个例子，掷一颗色子，问“掷出偶数或掷出大于 4 的点数”的可能性是多少？

如图 8 所示，定义结果为偶数对应事件 A ， $A = \{2, 4, 6\}$ ；点数大于 4 对应事件 B ， $B = \{5, 6\}$ 。

? 请大家绘制图 8 的韦恩图。

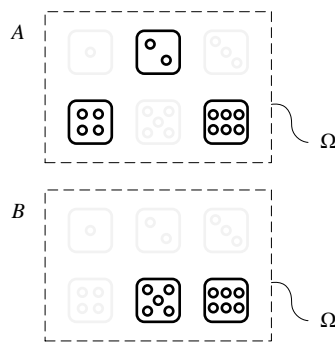


图 8. 事件 A 代表结果为偶数，事件 B 代表点数大于 4

比较图 8 中 A 、 B 事件，我们发现 6 重复计算，所以为了计算 $A \cup B$ 要减去重复的结果，即 $A \cap B = \{6\}$ ；因此，

$$A \cup B = \{2, 4, 5, 6\} \quad (7)$$

假设样本空间 $\Omega = \{1, 2, 3, 4, 5, 6\}$ 中每个样本等可能，则有

$$\begin{aligned} \Pr(A) &= \frac{3}{6} \\ \Pr(B) &= \frac{2}{6} \\ \Pr(A \cap B) &= \frac{1}{6} \end{aligned} \quad (8)$$

因此，

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) = \frac{3}{6} + \frac{2}{6} - \frac{1}{6} = \frac{2}{3} \quad (9)$$

互斥事件

如果事件 A 、 B 互斥，这意味着 A 和 B 不可能同时发生，即

$$A \cap B = \emptyset \Rightarrow \Pr \left(\underset{\text{Joint}}{A \cap B} \right) = \Pr \left(\underset{\text{Joint}}{A, B} \right) = 0 \quad (10)$$

如果 A 、 B 互斥，那么“发生 A 或发生 B ”的概率就等于它们各自概率的和，即

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) \quad (11)$$

如果事件 A_1 、 A_2 、 $A_3 \dots A_n$ 互不相容，则

$$\Pr(A_1 + A_2 + \dots + A_n) = \Pr(A_1) + \Pr(A_2) + \dots + \Pr(A_n) \quad (12)$$

意思是，如果多个事件两两互不相容，那么它们“或”发生的概率就是它们各自概率之和。


举个例子，如图9所示，掷一颗色子，事件 A 为“掷出的点数小于3”， $A = \{1, 2\}$ ；定义事件 B 为“掷出点数大于4”， $B = \{5, 6\}$ 。

根据等概率原理，

$$\begin{aligned} \Pr(A) &= \frac{2}{6} \\ \Pr(B) &= \frac{2}{6} \end{aligned} \quad (13)$$

显然事件 A 、 B 互斥， $A \cup B = \{1, 2, 5, 6\}$ 。色子点数“小于3或大于4”的概率为

$$\Pr(A \cup B) = \frac{4}{6} = \frac{1}{3} \quad (14)$$

 请大家绘制图9的韦恩图。

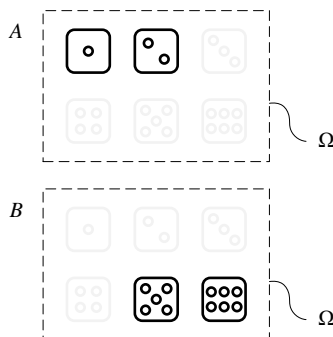


图9. 事件 A 为结果点数小于3，事件 B 为结果点数大于4

对立事件

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>


本书配套微视频均发布在 B 站——生姜 DrGinger：<https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：jiang.visualize.ml@gmail.com

事件 A 与其对立事件 \bar{A} 必有一个发生，因此两者概率之和为 1，即

$$\Pr(A + \bar{A}) = \Pr(\Omega) = 1 \quad (15)$$

比如，如图 10 所示，掷一颗色子，事件 A 代表结果为奇数，即 $A = \{1, 3, 5\}$ ；事件 B 代表结果为偶数， $B = \{2, 4, 6\}$ 。

 请大家计算图 10 中事件 A 、 B 的概率。

一方面，和事件 $A \cup B$ 为整个样本空间，即 $A \cup B = \{1, 2, 3, 4, 5, 6\} = \Omega$ ；另一方面， A 、 B 互斥，即 $A \cap B = \{\} = \emptyset$ 。显然， A 、 B 为对立事件。

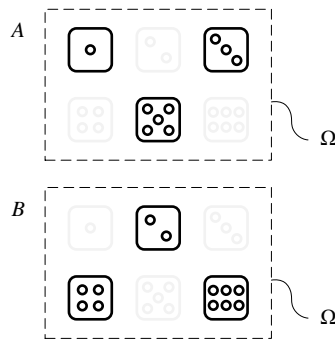


图 10. 事件 A 代表结果为奇数，事件 B 代表结果为偶数



请大家用 DeepSeek/ChatGPT 等工具完成本节如下习题。

Q1. 抛两枚均匀硬币，计算如下事件概率

- ▶ 两枚硬币均为正面；
- ▶ 至少一枚硬币为正面；
- ▶ 两枚硬币均为反面；

Q2. 掷两颗色子，计算如下事件概率

- ▶ 第一、二颗色子点数都为 6；
- ▶ 第一颗色子点数为 6；
- ▶ 第二颗色子点数为偶数；
- ▶ 第一、二颗色子点数之和为 8；
- ▶ 第一颗色子点数比第二颗大 2；
- ▶ 两颗色子点数之差为 2；
- ▶ 两颗色子点数之和不小于 8；
- ▶ 两颗色子点数之和小于 8；

本 PDF 文件为作者草稿，发布目的为方便读者在移动终端学习，终稿内容以清华大学出版社纸质出版物为准。

版权归清华大学出版社所有，请勿商用，引用请注明出处。

代码及 PDF 文件下载：<https://github.com/Visualize-ML>

本书配套微视频均发布在 B 站——生姜 DrGinger：<https://space.bilibili.com/513194466>

欢迎大家批评指教，本书专属邮箱：jiang.visualize.ml@gmail.com

► 两颗色子取值均在 $[2, 6]$ 区间；

Q3. 编写 Python 代码，进行随机试验：每次同时掷两颗色子，重复试验 100 次，估计“点数之和为 8”的概率。绘制概率随试验次数变化图像。

Q4. 什么是“赌徒谬误”？