

MAT220 第8讲讲义

二项分布 第1部分

目录

1	简短复习：计数与分布	2
1.1	计算3次抛硬币中恰好出现2次反面的概率	2
1.2	这个问题的一般解法	3
1.3	概率分布	3
1.4	大数定律	5
2	例子：二项实验	7
2.1	例1：抛一枚公平硬币10次	7
2.2	其他 n 值下的二项分布	8
2.3	例2：抛一枚不公平硬币	9
3	一般理论	10
3.1	二项实验的定义	10
3.2	p 会如何影响分布？	11
3.3	二项分布的期望值和方差	12
4	练习	12
4.1	轮盘	12
4.2	抽卡箱	14
4.3	回转寿司	15
4.4	掷骰子	16
4.5	打篮球的猫	17

在本讲中，我们讨论概率论中的最后几个主题之一：二项实验。正如我们将看到的，这是一类特殊的设置，其中只有两种可能的结果，即“成功”或“失败”。由于只有两种可能性，这类系统相当容易描述和计算，这意味着我们可以很容易地把它们与之前的内容联系起来。在介绍一般理论之后，我们将讨论二项实验的几个例子。

1 简短复习：计数与分布

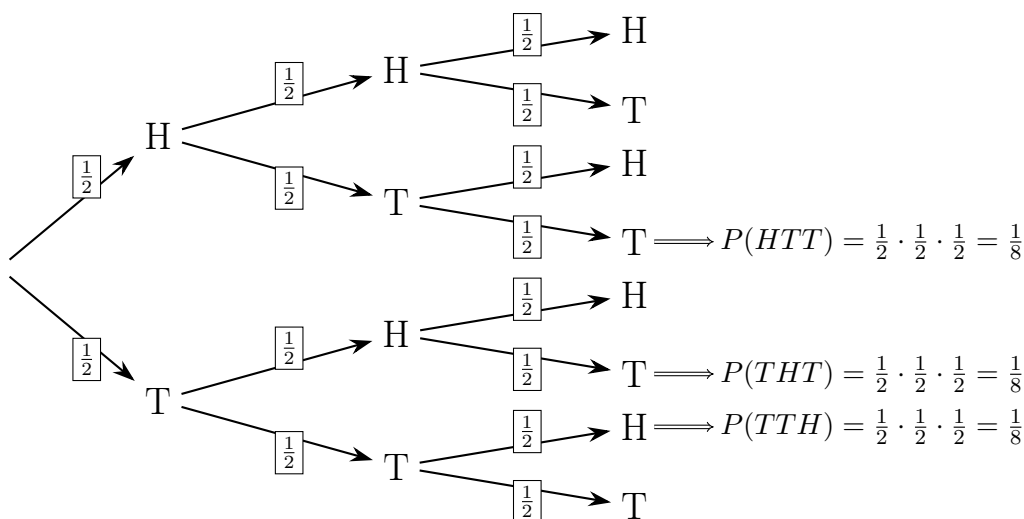
1.1 计算3次抛硬币中恰好出现2次反面的概率

假设我们有一枚公平的硬币，并且抛这枚硬币3次。恰好得到2次反面的概率是多少？幸运的是，像3这样的小数，我们可以直接写出所有可能的结果：

- 一次抛硬币有2种结果： H 或 T 。
- 两次抛硬币有4种结果： HH, HT, TH, TT 。
- 三次抛硬币有8种结果（完整的样本空间）：

$$\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

我们有几种方法可以计算像 $P(\text{恰好 } 2 \text{ 个 } T)$ 这样的概率。一种方法是使用决策树（类似第7讲中的方法）来计算每一个有利结果的概率：



HTT , TTH 和 THT 这三个事件互斥，所以我们可以直接把这些单个结果的概率相加，这就给出了这个复合事件的概率：

$$P(\text{恰好 } 2 \text{ 个 } T) = P(HTT \text{ 或 } THT \text{ 或 } TTH) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{3}{8}.$$

另一种方法是用更简单的复合事件计算来得到同一个答案（类似第5讲）。我们可以观察到，在这个情形中，所有结果都是等可能的，所以也可以使用黄金公式：

$$P(\text{事件}) = \frac{\text{有利于该事件的结果数}}{\text{结果总数}}.$$

在这个例子中，恰好包含两个反面的结果有3个，即 HTT , THT 和 TTH ，所以：

$$P(\text{恰好 } 2 \text{ 个 } T) = \frac{\text{包含 } 2 \text{ 个 } T \text{ 的结果数}}{\text{结果总数}} = \frac{3}{8}.$$

1.2 这个问题的一般解法

如果现在我们抛硬币 n 次，那么样本空间会有 2^n 个结果。为了求“恰好 r 次反面”的概率，我们想计算如下比值：

$$\frac{\text{恰好}r\text{次反面的结果数}}{2^n}.$$

因此，只要我们找到一种方法来数出 n 次抛硬币中恰好 r 次反面的结果数，这个问题就可以解决。

一种理解方法如下：把 n 次抛硬币的结果描述为由H和T组成的字符串。例如，5次抛硬币的一个字符串可能是 $HHTHT$ 。在我们真正抛硬币之前，可以想象这些位置都是空白的、还没有被指定。例如，当 $n = 5$ 时，我们可以把可能的结果想象成 $_ , _ , _ , _ , _$ 。现在，要求恰好 r 次反面，就相当于问有多少种方法可以把这些位置中的 r 个指定为T。一旦我们选择了哪 r 个位置是T，其余每个位置就都被迫是H，所以每一种T的位置选择都编码了一个唯一的结果。不同的位置选择给出不同的结果。因此，在 n 个可用位置中，我们必须选择 r 个位置来指定为T。换句话说，我们有 n 个潜在选择，并且必须恰好选定 r 个。这正是一个组合问题，也就是说，我们是在 n 个抛硬币位置中选择一个 r 元素子集，因此不同选择的数量由公式 $C_{n,r}$ 计算。

在第4讲中，我们看到，从大小为 n 的集合中选择 r 个元素的方法数，可以直接从帕斯卡三角形中读出：

[illegible]

具体来说，答案可以通过进入第 n 行（从第0行开始数），然后找到对应 r 的项（从 $r = 0$ 开始数）来得到。这就给出了 $C_{n,r}$ ，也就是我们想要的数。例如，在第三行中 $r = 2$ 对应的项是3，这对应于3次抛硬币中得到2次反面的方式数。

现在假设我们抛硬币5次。恰好得到2次反面的概率是多少？

解答

结果总数为 $2^5 = 32$ 。恰好有2次反面的结果数是 $C_{5,2} = 10$ （帕斯卡三角形第5行）。因此，

$$P(\text{恰好 2 次反面}) = \frac{C_{5,2}}{2^5} = \frac{10}{32} = \frac{5}{16}.$$

1.3 概率分布

如果变量 X 在某个实验或观察中所取的值由随机结果决定，那么 X 称为随机变量。主要有两种类型

型:

- 离散型随机变量只能取有限个值，或者可数多个值。
- 连续型随机变量可以取某个区间中的无数多个值中的任意一个。

练习：离散型还是连续型？

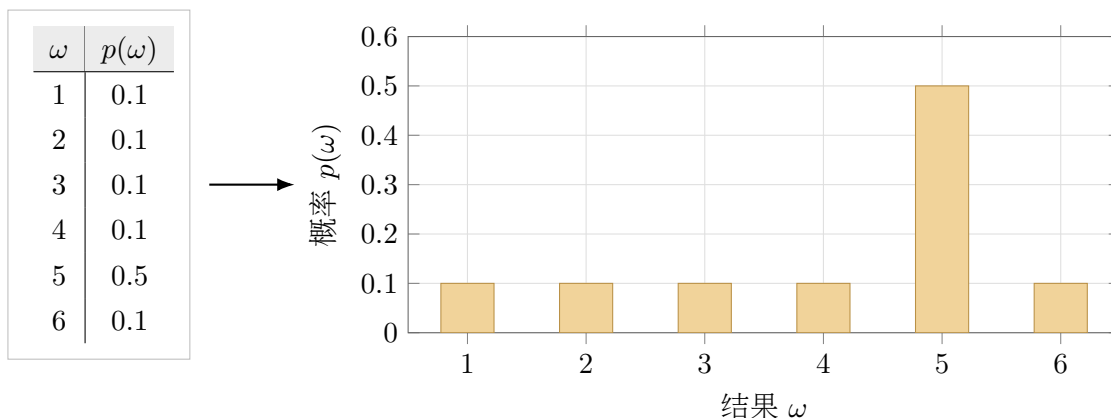
把每个随机变量分类为离散型或连续型。

- (a) 某个被选中的学生完成秋季学期注册所需的时间。
- (b) 某个被选中的学生在随机选定的一天收到的短信数量。
- (c) 一辆电动汽车满电后可以行驶的英里数。
- (d) 从某地区的登记选民中随机选取50人，并统计他们中有多少人在上一次县级选举中投票。

解答

- (a) 连续型。
- (b) 离散型。
- (c) 连续型。
- (d) 离散型。

离散概率分布是对离散型随机变量的每一个不同取值分配概率。在第6讲中，我们看到概率分布可以用柱状图可视化。例如，下图：



直观地说明了概率如何分布在可能的值1, 2, 3, 4, 5, 6之间。

通常我们用大写字母来表示随机变量，例如 X 或 Y 。如果 X 可以取值1, 2, 3, 4, 5, 6，那么像

$$X = 4 \quad \text{或} \quad X = 5$$

这样的陈述表示随机变量的可能取值。 X 取值4的概率可以写成：

$$P(X = 4).$$

在概率分布中，这由4处柱子的高度表示。在上面的例子中，

$$P(X = 4) = 0.1.$$

我们也可以询问 X 落在某个取值范围内的概率。例如：

$$P(X < 3)$$

表示 X 取小于3的值的概率。由于小于3的值是1和2，所以我们可以通过求和来计算总概率：

$$P(X < 3) = P(X = 1) + P(X = 2) = 0.1 + 0.1 = 0.2.$$

在第6讲中，我们还看到过随机变量的期望值公式，它类似于概率分布的平均数。概率系统的期望值是加权平均。对于有限离散型随机变量，期望值是

$$E(X) = \sum_{i=1}^n p_i x_i = p_1 x_1 + \cdots + p_n x_n.$$

这个公式可以通过尝试计算概率分布的几何质心来仔细推导。完整解释请参考第6讲。非正式地说， $E(X)$ 是 X 的长期期望行为。

概率分布的方差衡量 X 的可能取值围绕其期望值分散的程度。它定义为

$$\text{Var}(X) = E((X - E(X))^2).$$

注意这个公式与数据集的总体方差通常定义之间的相似性。总结如下：

期望值与方差

离散型随机变量的**期望值**是

$$E(X) = \sum_{i=1}^n p_i x_i.$$

它是平均数的概率版本，因此给出了分布中心的度量。

随机变量的**方差**是

$$\text{Var}(X) = E((X - E(X))^2).$$

它衡量分布围绕其期望值分散的程度。

1.4 大数定律

随机变量的期望值是一种计算实验平均的、长期期望结果的方法。现在我们通过引入概率的另一种解释来进一步说明这个思想。

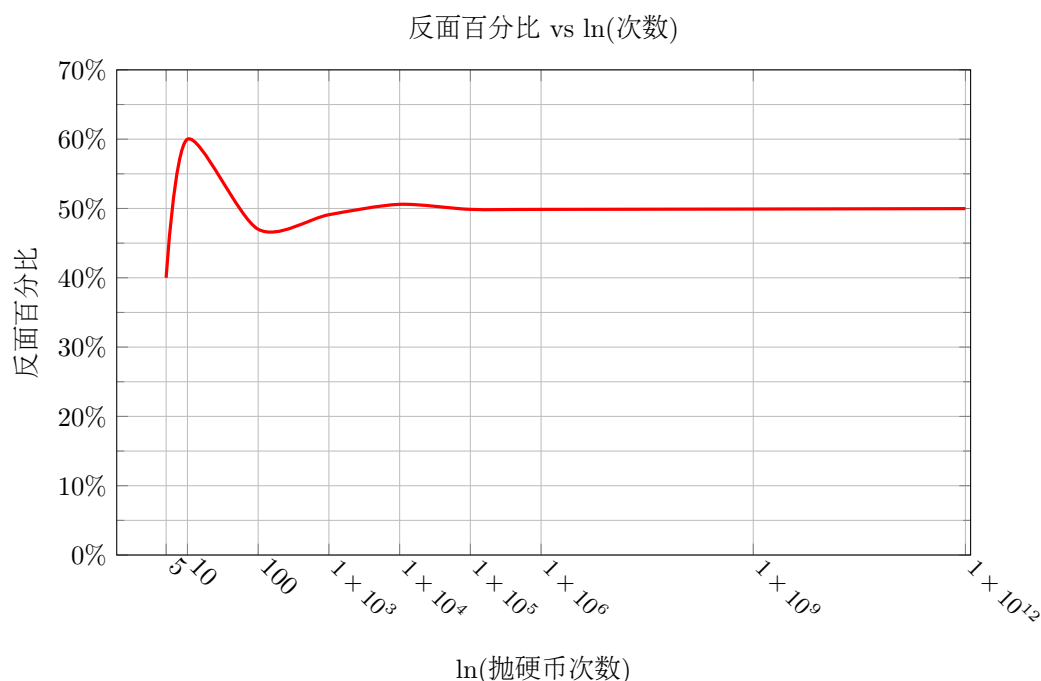
大数定律说明，如果你重复某个实验许多次，那么平均结果会趋向于接近理论平均。作为例子，假设我们有一枚公平硬币。我们预期抛硬币并得到反面的概率为50%，即 $P(T) = \frac{1}{2}$ 。我们的概率解释说明，从长期来看，我们预期大约50%的结果是反面。

为了看到这一点，假设我们开始一遍又一遍地抛硬币。使用随机数生成器，下面是硬币投掷样本中反面比例的一个例子：

- 5次：HHTTH。 $T = 2$, $H = 3$ ，所以反面 = 40%。

- 10次: HHTTHTHTTT。 $T = 6, H = 4$, 所以反面= 60%。
- 100次: $T = 47, H = 53$, 所以反面= 47%。
- 1000次: $T = 491, H = 509$, 所以反面= 49.1%。
- 10000次: $T = 5062, H = 4938$, 所以反面= 50.6%。
- 100000次: $T = 49872, H = 50128$, 所以反面= 49.87%。
- 1百万次: 反面= 49.87%。
- 1十亿次: 反面= 49.92%。
- 1万亿次: 反面= 49.982%。

我们可以把它画成图。这里使用对数尺度 $\ln(\cdot)$, 使得这些抛硬币次数可以放在同一条轴上:



如你所见, 开始时结果有相当大的随机波动。然而, 随着时间推移, 这些随机变化大多会相互抵消, 反面的比例收敛到理论平均值50%。这为我们提供了一种可能的概率解释:

概率的频率主义解释

当我们说 $P(A) = x\%$ 时, 我们的意思是, 如果我们把实验重复非常多次, 那么对应事件 A 的结果的长期比例会越来越接近 $x\%$ 。

从这个意义上说, 期望值 $E(X)$ 可以更好地理解: 在独立条件下不断重复该实验时会得到的平均结果。最终, 这种重复活动的运行平均会收敛到某个东西, 而这个东西就是期望值。正如我们稍后会看到的, 期望值在讨论二项实验时特别有用。

2 例子：二项实验

2.1 例1：抛一枚公平硬币10次

假设我们有一枚公平硬币，并且连续抛10次。恰好得到5次反面的概率是多少？

解答

根据上一节，我们可以用以下公式计算 $P(\text{恰好 } 5 \text{ 次反面})$ ：

$$P(\text{恰好 } 5 \text{ 次反面}) = \frac{C_{10,5}}{2^{10}}.$$

当然，我们也可以使用通常的公式 $C_{n,r} = \frac{n!}{r!(n-r)!}$ ，但为了叙述上的一致性，我们改用帕斯卡三角形。三角形的第10行是：

1 10 45 120 210 252 210 120 45 10 1

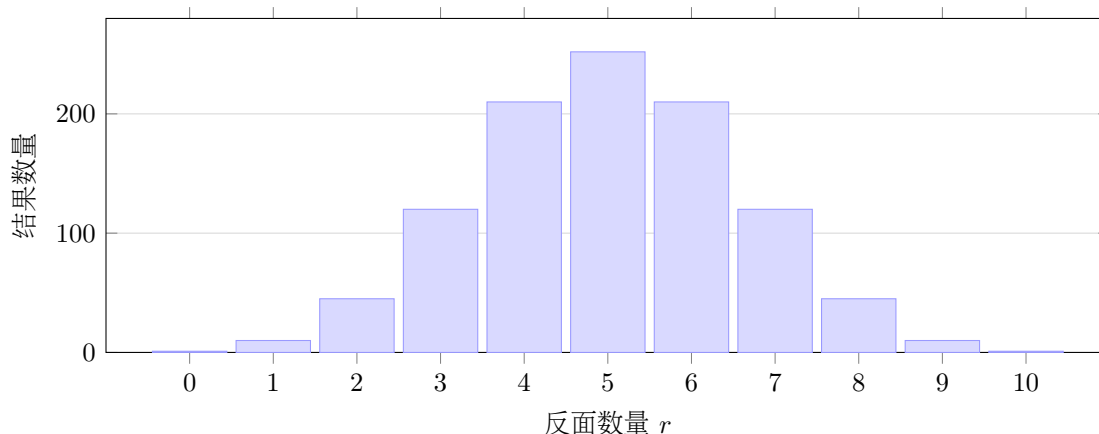
因此，我们移动到 $r = 5$ 对应的项（记住三角形从0开始计数），可以看到 $C_{10,5} = 252$ 。结果总数为 $2^{10} = 1024$ 。因此：

$$P(5 \text{ 次反面}) = \frac{C_{10,5}}{2^{10}} = \frac{252}{1024} \approx 0.246 \approx 24.6\%.$$

现在我们尝试研究10次抛硬币的样本空间结构。根据上面的讨论，我们知道这个空间有1024个成员。它们从 $HHHHHHHHHH$ 到 $TTTTTTTTTT$ ，并包含中间的所有结果。整理这个大空间的一种自然方式，是根据结果中包含多少个反面来分组。¹ 理解样本空间的一种有用方式是按结果中包含的反面数量来分组。根据帕斯卡三角形第10行，包含1024个结果的样本空间可以组织如下。

反面数量 r	0	1	2	3	4	5	6	7	8	9	10
结果数量	1	10	45	120	210	252	210	120	45	10	1

上面的表可以可视化为柱状图：



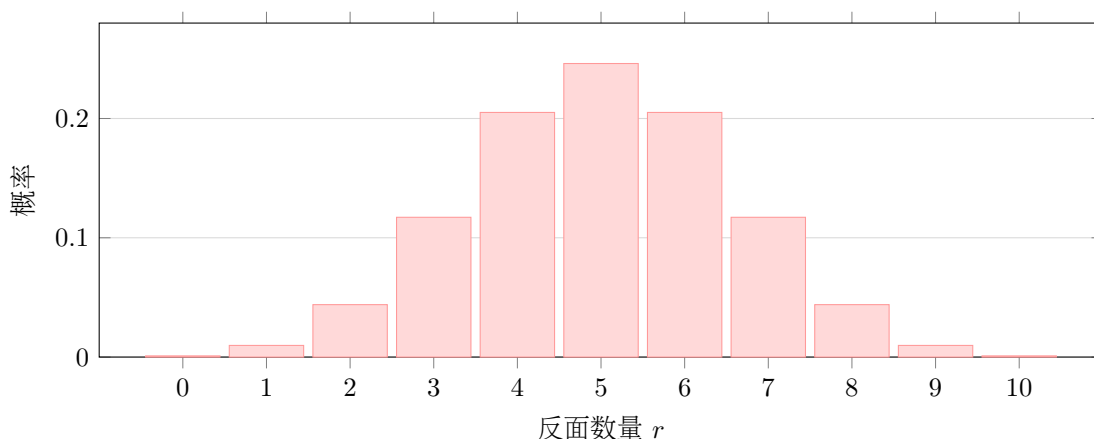
¹当然，也可以从正面开始做同样的事情。但是，由于我们一直在使用反面，所以这里继续使用反面。

这里 x 轴由一个离散计数组成：简单来说，它就是可能的反面数量。 y 轴是对应结果的数量。因此，在某种意义上，这个图只是帕斯卡三角形第10行的可视化。

根据上面的讨论，我们知道这个系统中的概率可以通过把 $C_{10,r}$ 除以 $2^{10} = 1024$ 来计算。因此，我们可以进行这个除法，然后把结果画出来，看看概率本身如何分布在这个系统的结果之间。

r 次反面	0	1	2	3	4	5	6	7	8	9	10
概率 (分数)	$\frac{1}{1024}$	$\frac{10}{1024}$	$\frac{45}{1024}$	$\frac{120}{1024}$	$\frac{210}{1024}$	$\frac{252}{1024}$	$\frac{210}{1024}$	$\frac{120}{1024}$	$\frac{45}{1024}$	$\frac{10}{1024}$	$\frac{1}{1024}$
近似百分比	0.1%	1.0%	4.4%	11.7%	20.5%	24.6%	20.5%	11.7%	4.4%	1.0%	0.1%

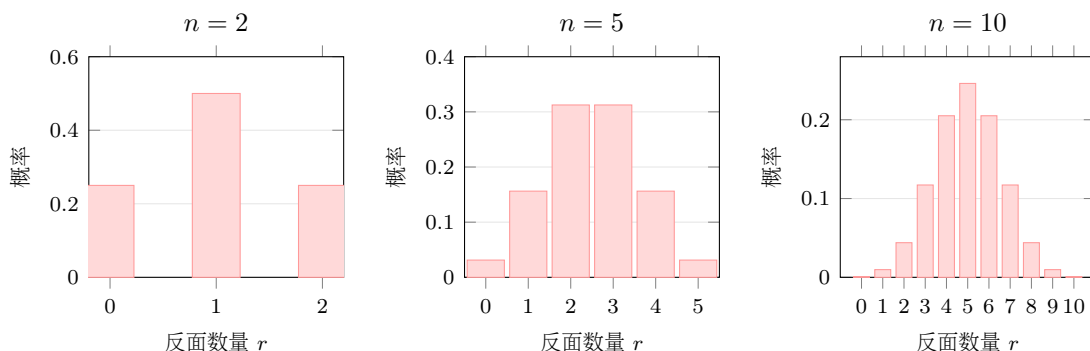
注意我们之前得到的24.6%这个值。使用上面的表，我们可以画出这个概率分布：

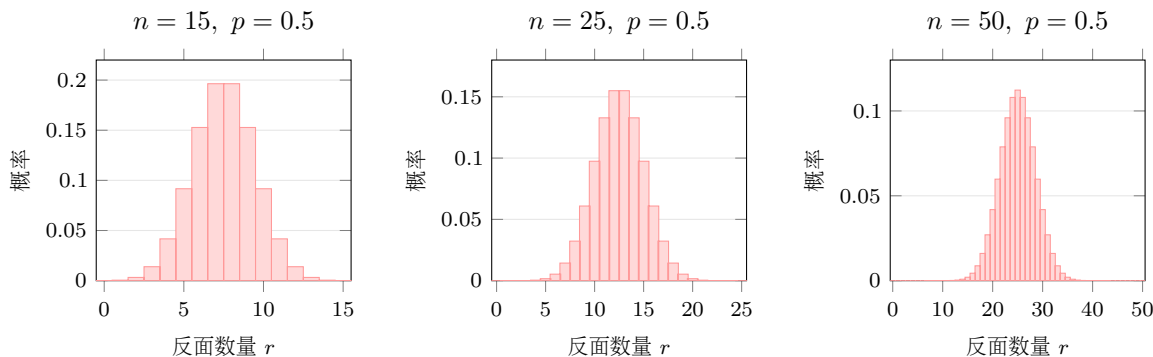


请注意，它的形状与前面的柱状图完全相同。唯一改变的是 y 轴的尺度。这个特殊的图有一个名字：它叫做二项分布。

2.2 其他 n 值下的二项分布

一般来说，我们可以为 n 的其他取值创建二项分布，其中 n 是抛硬币的总次数。这些图的形状会展示概率如何分布在抛硬币的大量可能结果之间。这个形状来自帕斯卡三角形的第 n 行，把这些数除以 2^n 之后，它们就变成了概率。一般来说，如果我们继续增加抛硬币次数（也就是让 n 变大），那么 r 的可能取值会更多，分布会变得更平滑，并且会更加集中在中间附近（接近 $r = n/2$ ）。我们会在第10讲中详细讨论这种“平滑性”。





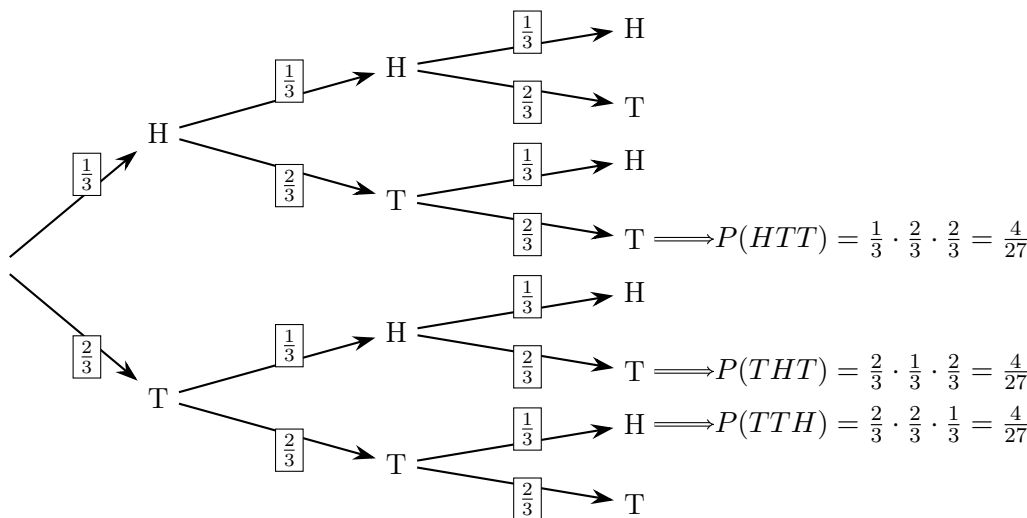
2.3 例2：抛一枚不公平硬币

现在假设我们有一枚以某种方式被动过手脚的硬币，因此正面或反面的概率不再相等。我们假设这枚“不公平”的硬币有如下概率：

$$P(H) = \frac{1}{3}, \quad P(T) = \frac{2}{3}.$$

问题：抛3次时，恰好得到2次反面的概率是多少？

在上面讨论的公平硬币情形中，所有8个结果都是等可能的，所以我们可以直接数出所有有利结果（有3个），然后除以8。然而，对于这枚不公平硬币，并不是所有结果都是等可能的。因此，我们不能只数有利结果再除以8。相反，我们可以使用决策树方法。可能结果的空间与之前相同，但是现在树的每条分支上的概率发生了变化：



恰好2次反面对应于结果HTT, THT, TTH。这些是互斥事件，所以我们可以通过求和得到所需概率：

$$P(\text{恰好 2 次反面}) = P(HTT \text{ 或 } THT \text{ 或 } TTH) = \frac{4}{27} + \frac{4}{27} + \frac{4}{27} = \frac{12}{27}.$$

注意，每一个结果的概率计算本质上是相同的：为了计算 $P(HTT)$ 、 $P(THT)$ 或 $P(TTH)$ ，我们只需要把两个 $P(T) = \frac{2}{3}$ 和一个 $P(H) = \frac{1}{3}$ 相乘。事实上，这种计算方式也适用于进一步的概率计算。对于 n 次抛掷，我们有：

$$P(r \text{ 个 } T) = C_{n,r} \left(\frac{2}{3}\right)^r \left(\frac{1}{3}\right)^{n-r}.$$

更一般地，如果我们把得到反面或正面的概率调整为其他值，那么 n 次抛硬币中得到 r 次反面的概率可以用公式

$$P(n \text{ 次抛掷中 } r \text{ 个 } T) = C_{n,r} (P(T))^r (P(H))^{n-r}.$$

计算。这里 $(P(T))^r (P(H))^{n-r}$ 是包含 r 个 T 和 $(n-r)$ 个 H 的某一个特定顺序的概率，而 $C_{n,r}$ 计算的是 r 个 T 和 $(n-r)$ 个 H 的不同排列顺序数量。

3 一般理论

3.1 二项实验的定义

在第2节中，我们看到了两种不同版本的抛硬币：一种是公平的，另一种在两个可能结果具有不同概率的意义上是“不公平”的。事实上，这两个都是一种更一般过程的例子，这种过程叫作二项实验。二项实验有四个重要的定义性特征。

定义：二项实验

二项实验具有以下性质：

- (B1) 有固定数量的试行。把这个数量记作 n 。
- (B2) 这 n 次试行相互独立，并且在相同条件下重复进行。
- (B3) 每次试行只有两种结果：成功（ S ）和失败（ F ）。
- (B4) 每次试行中成功的概率相同。把它记作 p ，把失败的概率记作 q 。由于每次试行不是成功就是失败，所以 $p + q = 1$ ，因此 $q = 1 - p$ 。

换句话说，二项实验就是固定次数的独立试行，每次试行只有两种可能结果，并且成功概率相同。在二项实验中，核心问题是确定 n 次试行中 r 次成功的概率。假设(B2)用于简化这些连续概率的计算。

3.1.1 例子：公平硬币投掷

在前一节的第一个例子中：

- (B1) n 是抛硬币的次数。
- (B2) 抛硬币之间是独立的。事实上，它们与之前实验的结果完全无关（甚至在物理上也无关）。
- (B3) 成功被定义为“得到反面”，失败被定义为“没有得到反面”。
- (B4) 对于公平硬币， $p = P(T) = 0.5$ 并且 $q = 0.5$ 。

对于确定这个实验相关概率的核心问题，我们之前的讨论正是在做这件事。我们已经看到，对于公平硬币，

$$P(n \text{ 次抛掷中 } r \text{ 次反面}) = \frac{C_{n,r}}{2^n}$$

成立。

对于不公平硬币，二项设置本质上相同，只是我们把概率调整为 $p = \frac{2}{3}$ 和 $q = \frac{1}{3}$ 。我们之前看到，对于不同的 r 值，相关概率可以用下面的公式计算：

$$P(r \uparrow T) = C_{n,r} \left(\frac{2}{3}\right)^r \left(\frac{1}{3}\right)^{n-r} = C_{n,r} \cdot P(T)^r \cdot P(H)^{n-r}.$$

二项概率的一般公式

对于参数为 $n, p, q = 1 - p$ 的二项实验，恰好得到 r 次成功的概率为

$$P(n \text{ 次试行中 } r \text{ 次成功}) = C_{n,r} p^r q^{n-r}.$$

这里：

- p^r 是某一个特定顺序的概率中的成功部分。
- q^{n-r} 是同一个特定顺序的概率中的失败部分。
- $p^r q^{n-r}$ 是一个包含 r 次成功和 $(n - r)$ 次失败的特定序列的概率。
- $C_{n,r}$ 计算有多少个不同序列包含 r 次成功和 $(n - r)$ 次失败。

3.2 p 会如何影响分布？

一般来说，我们可以用公式

$$P(n \text{ 次试行中 } r \text{ 次成功}) = C_{n,r} p^r q^{n-r}.$$

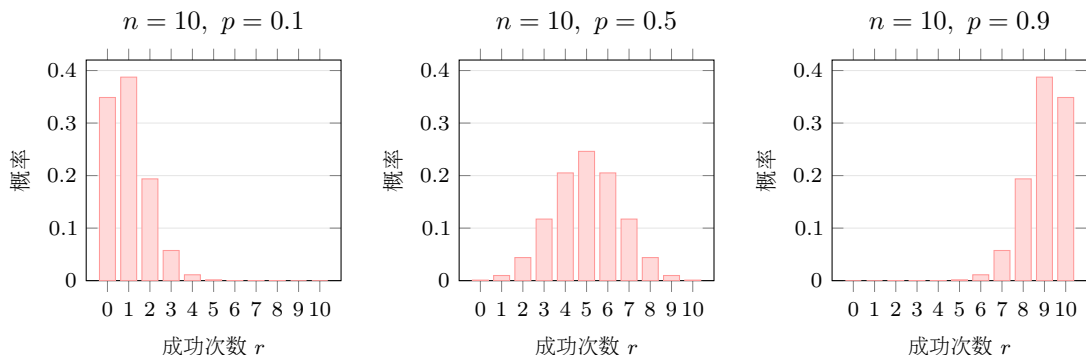
来描述每一个二项实验所对应的二项分布。简单来说，我们把 x 轴设为随机变量的可能取值（从0到 n ），把 y 轴设为上面公式的输出。

在公平硬币的情形中，成功概率 $p = 0.5$ ，这正好等于 q 。这会使得到的二项分布完全对称。请在第2节中的图中注意这一点。如果我们开始把成功概率从 $p = 0.5$ 调整开，那么分布就会失去对称性。

为了说明这一点，假设我们正在研究 $n = 10$ 次试行的二项实验。当概率 $p = 0.5$ 时，得到的分布就是第2.1节中的图。现在我们简单讨论如果让 p 变大或变小会发生什么。

- 如果 p 接近0，例如 $p = 0.1$ ，那么每次试行中得到成功结果的机会就低得多。因此，在10次试行中得到大量成功，例如7, 8, 9或10次成功，是极其不可能的。最可能的是，在10次尝试中只得到0, 1或2次成功。因此，与这种系统相关的二项分布会把大部分概率分布在 x 轴的左侧，靠近0, 1, 2这些值。
- 如果 p 接近1，例如 $p = 0.9$ ，那么每次试行得到成功就会极其可能。因此，在10次尝试中只有一两次成功会非常不走运，而你反而会期望得到大约7, 8, 9或10次成功。在二项分布层面，这样高的成功概率会把分布推向右侧，并使大部分概率集中在 x 轴的8, 9, 10附近。

下面的图分别展示了 $n = 10$ 次试行，成功概率为0.1, 0.5, 0.9时的实际二项分布。



3.3 二项分布的期望值和方差

对于二项分布，有两个特别重要的公式：

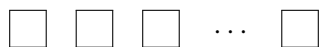
二项期望值和方差

如果 X 表示参数为 n , p 和 $q = 1 - p$ 的二项实验中的成功次数，那么

$$E(X) = np, \quad \text{Var}(X) = npq.$$

这些公式很有用，因为它们让我们无需手动计算整个分布，就可以快速理解二项分布的中心和分散程度。

从技术上讲，我们可以用一些代数来验证上面的两个公式。不过，这里我们不这样做，而是给出一个粗略的解释：把 n 次试行想象成 n 个槽位：



每个槽位贡献1次成功或0次成功。由于每个槽位成功的概率是 p ，根据概率的频率主义解释，每个槽位在长期平均意义下贡献 p 次成功。由于有 n 个槽位，总的期望成功次数就是

$$p + p + \cdots + p = np,$$

这就得到公式 $E(X) = np$ 。至于方差公式，我们这里不推导。不过，我们要注意公式

$$\text{Var}(X) = npq$$

告诉我们成功次数的波动程度：

- 如果 p 接近0或接近1，那么实验波动较小，并且更紧密地集中在 $E(X)$ 附近，因为结果更容易预测。
- 如果 p 接近0.5，那么实验波动更大，因为每次试行中成功和失败都相当可能。

4 练习

4.1 轮盘

假设我们班有20名学生。你努力学习，几年后顺利从LUJ毕业。Dr. O'Connell为你感到非常骄傲，于是决定带你们所有人去赌场玩轮盘。每名同学转一次轮盘。



轮盘上有36个编号槽，绿色槽使得总结果数为37。Dr. O'Connell给出如下提议：

如果球落在绿色槽上，你得到\$50,000；如果出现任何其他结果，你得到\$0。

我们要回答的主要问题是：

练习

恰好3名学生获胜的概率是多少？

为了解决这个问题，我们将使用第3.2节中提到的一般二项公式。不过，在开始计算任何东西之前，我们必须先确认这个情形确实是一个二项实验。

- B1: 有20名学生，每人尝试一次轮盘，所以 $n = 20$ 次试行。
- B2: 假设每次转动都是独立的，并且轮盘是公平的，那么这些试行确实彼此独立。
- B3: 每次试行只有两种可能结果：

成功 = 落在绿色槽上， 失败 = 落在其他地方。

- B4: 轮盘上的每个槽都有相同的机会成为结果。由于轮盘上有37个槽，并且只有一个绿色槽可以落入，成功和失败的概率为：

$$p = P(\text{成功}) = \frac{1}{37}, \quad q = P(\text{失败}) = 1 - p = \frac{36}{37}.$$

现在我们已经确认这个系统可以看成二项实验，所以可以解主要问题。

解答

我们使用二项公式：

$$P(n\text{次试行中}r\text{次成功}) = C_{n,r} p^r q^{n-r}.$$

在这里，由于有20名学生，所以 $n = 20$ ， $r = 3$ ， $p = \frac{1}{37}$ ， $q = \frac{36}{37}$ 。把这些值代入公式，得到：

$$P(\text{恰好 3 人获胜}) = C_{20,3} \left(\frac{1}{37}\right)^3 \left(\frac{36}{37}\right)^{17} = \frac{20!}{3! \cdot 17!} \left(\frac{1}{37}\right)^3 \left(\frac{36}{37}\right)^{17}.$$

数值上，这大约是

$$P(\text{恰好 3 人获胜}) \approx 0.014 \approx 1.4\%.$$

作为这个概率的另一种解释：我们正在进行20次独立试行，并且想要3次成功和17次失败。能够表示这种情况的一个特定结果是：

$$S, S, S, F, F, \dots, F,$$

也就是说，前3名学生成功，之后17名学生全部失败。根据假设(B2)，所有这些试行彼此独立。因此，与这个结果相关的概率可以通过简单相乘每个结果的概率来计算。我们有：

$$P(S \text{ 然后 } S \text{ 然后 } S \text{ 然后 } F \text{ 然后 } \dots \text{ 然后 } F) = \frac{1}{37} \cdot \frac{1}{37} \cdot \frac{1}{37} \cdot \frac{36}{37} \cdot \dots \cdot \frac{36}{37},$$

这与写成 $p^3 q^{17}$ 相同。这是计算一个对应于3次成功和17次失败的单个结果的公式。但是，在20次试行中放置3次成功有 $C_{20,3}$ 种不同方式。这些序列对应互斥结果，所以我们把它们的概率相加，得到 $C_{20,3} p^3 q^{17}$ 。

$$P(\text{恰好 3 人获胜}) = \underbrace{p^3 q^{17} + p^3 q^{17} + \dots + p^3 q^{17}}_{C_{20,3} \text{ 次}} = C_{20,3} \cdot p^3 q^{17},$$

这与之前的公式相同。

4.2 抽卡箱

一个游戏中的稀有物品在每次游玩中出现的概率为0.26%。每次游玩花费\$2.49。我们把“成功”定义为获得该物品。

问题

- (a) 100次尝试中恰好获得2次成功的概率是多少？
- (b) 1000次尝试中恰好获得3次成功的概率是多少？
- (c) 在找到该物品之前，你应该期望支付多少钱？

解答

成功概率为

$$p = 0.0026, \quad q = 1 - p = 0.9974.$$

(a) 这里 $n = 100$ 且 $r = 2$ 。所以:

$$P(X = 2) = C_{100,2}(0.0026)^2(0.9974)^{98}.$$

数值上, 这大约是

$$P(X = 2) \approx 0.0259 = 2.59\%.$$

(b) 这里 $n = 1000$ 且 $r = 3$ 。所以:

$$P(X = 3) = C_{1000,3}(0.0026)^3(0.9974)^{997}.$$

数值上, 这大约是

$$P(X = 3) \approx 0.218 = 21.8\%.$$

(c) 我们要找的是使期望成功次数达到1的尝试次数 n 。由于 $E(X) = np$, 我们解:

$$1 = n(0.0026) \quad \implies \quad n = \frac{1}{0.0026} \approx 384.6.$$

因此, 平均而言, 我们会期望大约需要385次游玩才能得到一个稀有物品。由于每次游玩花费\$2.49, 这将花费

$$385 \times 2.49 = \$958.65.$$

4.3 回转寿司

一家回转寿司店允许你在把5个盘子放入机器后玩一次游戏并赢取奖品。赢得奖品的概率是25%。

问题

- (a) 玩5次并赢2次的概率是多少?
- (b) 2次尝试中赢2次的概率是多少?
- (c) 平均来说, 在赢得一个奖品之前, 你应该期望吃多少个盘子?

解答

成功定义为赢得奖品。那么

$$p = 0.25, \quad q = 0.75.$$

(a) 这里 $n = 5$ 且 $r = 2$, 所以:

$$P(X = 2) = C_{5,2}(0.25)^2(0.75)^3.$$

数值上,

$$P(X = 2) = 10(0.0625)(0.421875) = 0.263671875 \approx 26.4\%.$$

(b) 2次尝试中赢2次意味着两次都赢。由于尝试是独立的:

$$P(\text{赢然后赢}) = 0.25 \times 0.25 = 0.0625 = 6.25\%.$$

(c)

(d) 由于 $E(X) = np$, 我们希望 $E(X) = 1$, 所以:

$$1 = n(0.25) \quad \implies \quad n = 4.$$

因此, 我们期望大约4次尝试后赢得一个奖品。由于每次尝试需要5个盘子, 这对应于平均

$$4 \times 5 = 20$$

个盘子。

4.4 掷骰子

想象你掷一枚标准的6面骰子三次。

我们把“成功”定义为掷出6。那么每次掷骰只有两种结果（6或不是6），每次掷骰相互独立，并且成功概率恒定：

$$n = 3, \quad p = P(\text{掷出 } 6) = \frac{1}{6}, \quad q = \frac{5}{6}.$$

问题

恰好掷出两个6的概率是多少？

解答

要恰好掷出两个6，我们需要在 $n = 3$ 次试行中有 $r = 2$ 次成功。二项公式给出：

$$P(r \text{ 次成功}) = C_{n,r} p^r q^{n-r}.$$

所以

$$P(\text{恰好两个 } 6) = C_{3,2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right) = 3 \cdot \frac{1}{36} \cdot \frac{5}{6} = \frac{15}{216} = \frac{5}{72} \approx 6.9\%.$$

叙述解释：在三次掷骰中选择哪两次是6有 $C_{3,2} = 3$ 种方式，而每一种这样的模式的概率都是 $\left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)$ 。

4.5 打篮球的猫

一只猫正在投篮罚球。猫并不太擅长篮球，所以假设投进的概率是 $p = 0.5$ 。这只猫投6次。我们把“成功”定义为“投篮命中”。那么：

$$n = 6, \quad p = 0.5, \quad q = 0.5.$$

问题

- (a) 猫恰好投进4球的概率是多少？
- (b) 猫至少投进4球的概率是多少？

解答

令 X 为投进的球数。

- (a) 恰好投进4球意味着在 $n = 6$ 次试行中有 $r = 4$ 次成功：

$$P(X = 4) = C_{6,4} (0.5)^4 (0.5)^2 = C_{6,4} (0.5)^6 = \frac{C_{6,4}}{64} = \frac{15}{64} \approx 23.4\%.$$

- (b) “至少4球”意味着4球、5球或6球，所以我们把这些互斥情况相加：

$$\begin{aligned} P(X \geq 4) &= P(X = 4) + P(X = 5) + P(X = 6) \\ &= \frac{C_{6,4}}{64} + \frac{C_{6,5}}{64} + \frac{C_{6,6}}{64} = \frac{15 + 6 + 1}{64} = \frac{22}{64} = \frac{11}{32} \approx 34.4\%. \end{aligned}$$