

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

2.6 二项分布



本节你将掌握的核心技能：

- ▶ 二项分布描述的是重复 n 次独立试验后成功次数的分布；
- ▶ 从树状路径中数出每种结果出现的路径数量，并将其转化为概率；
- ▶ 参数 p 控制成功倾向， p 大时分布向右偏， p 小时向左偏；
- ▶ 随着试验次数增多，分布形状更集中、更平滑，并逐渐接近钟形。

连续抛硬币

二项分布 (binomial distribution) 是建立在伯努利分布基础上的一种重要离散概率分布，它描述的是：当我们对同一随机试验独立重复进行 n 次，每次试验只有“成功”或“失败”两种结果，而且成功的概率始终为同一个常数 p 时，最终观察到“成功”出现的总次数会呈现怎样的统计规律。

为了让这个概念更直观，可以继续沿用抛硬币的场景。假设我们有一枚硬币，正面朝上的概率是 p ，反面朝上的概率是 $1 - p$ ，并且每一次抛掷互不影响。

若把“正面出现”视为“成功”，那么独立抛 n 次硬币之后，正面出现的次数就是一个随机变量，我们通常记为 X 。二项分布研究的正是这个随机变量的分布规律：它告诉我们，出现 0 次正面、1 次正面、2 次正面……直到 n 次正面的概率分别是多少。

图 1 中展示的是对硬币进行三次独立抛掷的所有可能结果。每一次分叉代表一次抛硬币：向上的分支是正面 (head, H)，向下的分支是反面 (tail, T)。

从“Start”节点出发，硬币第一次被抛出，于是产生两个分支；随后每个节点再次分成两个分支，代表第二次和第三次抛掷的结果。

如果我们沿着图中的某一条路径从左读到右，就得到一次完整的三次抛硬币的结果序列，例如：Head → Head → Tail；Tail → Head → Head；Tail → Tail → Tail。

这幅图完整地列出了所有 $2^3 = 8$ 条路径，每条路径代表三次抛硬币的一个可能结果。

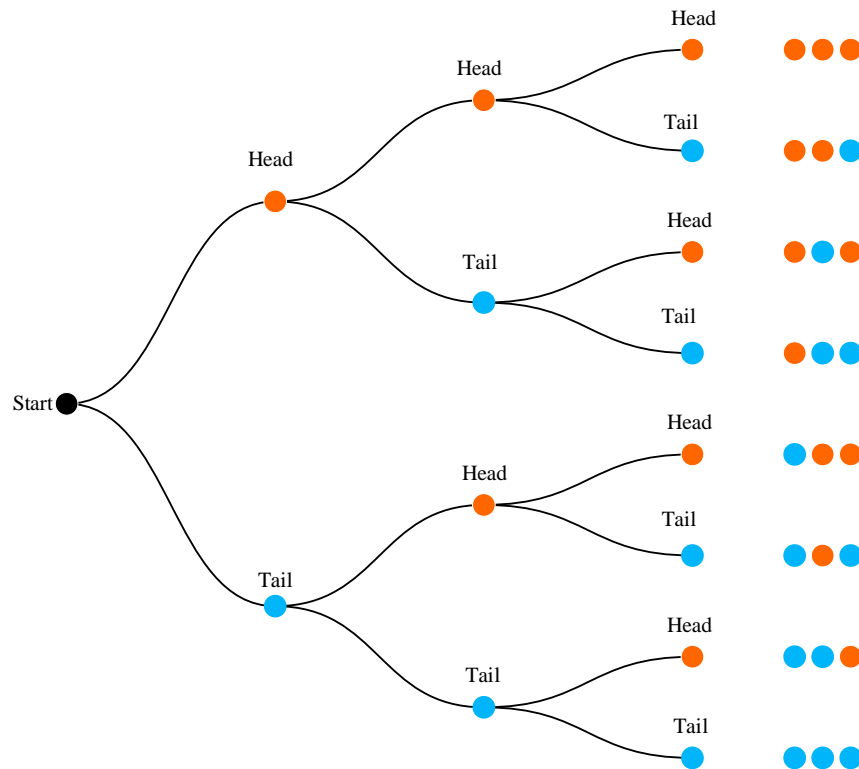


图 1. 一枚硬币抛 3 次，8 条路径

二项分布关心的不是某一条路径本身，而是“三次抛掷中出现了多少个 Head”，以及对应的概率。

图 1 中我们可以直观看到：

- 出现 3 次 Head 的路径：1 条 (HHH)，如图 2 (a)；
- 出现 2 次 Head 的路径：3 条 (HHT、HTH、THH)，如图 2 (b)；
- 出现 1 次 Head 的路径：3 条 (HTT、THT、TTH)，如图 2 (c)；
- 出现 0 次 Head 的路径：1 条 (TTT)，如图 2 (d)。

于是，“出现 x 次 Head” 对应的是“有多少条路径恰好包含 x 个 Head”。

下面，让我们看看如何计算概率。

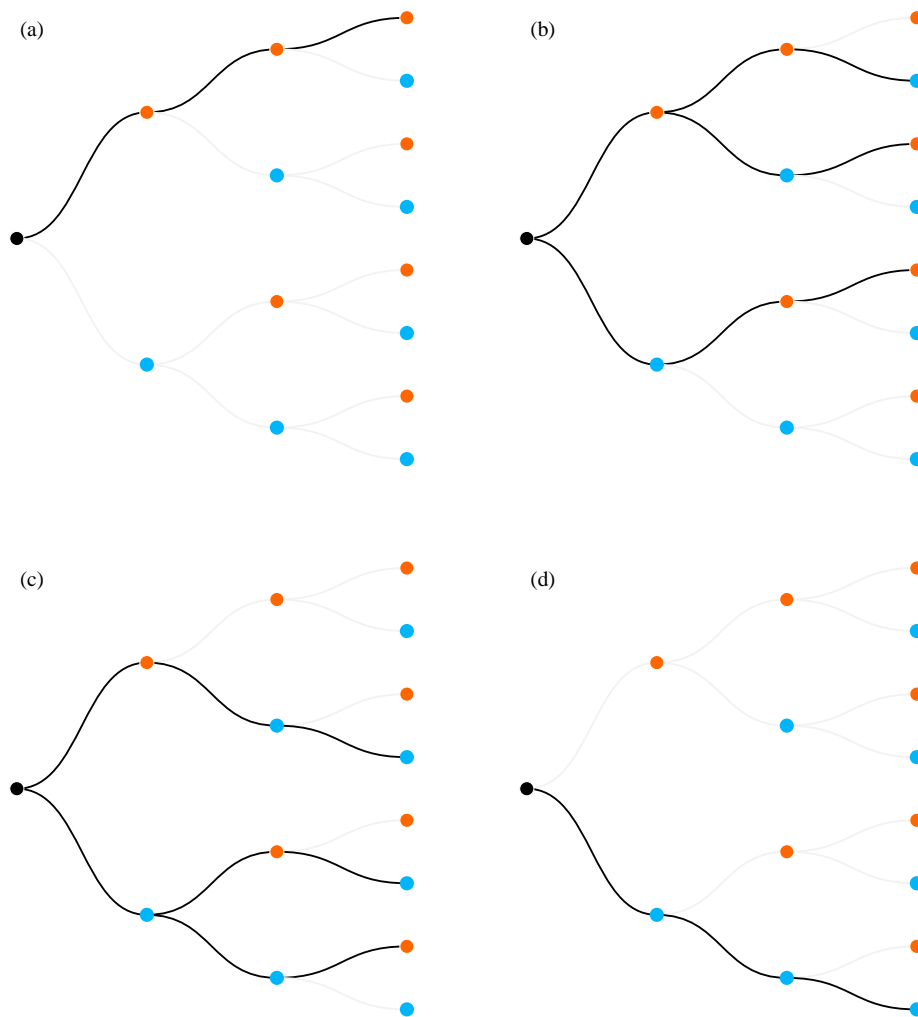


图 2. 一枚硬币抛 3 次，出现“正面”不同数量的路径

概率质量函数

有了上述讨论，我们发现在数学上，二项分布的关键在于：某一种具体的“成功次数”为 x 的结果，可以通过“有多少种方式在 n 次试验中选出 x 次成功”与“每一种方式本身的概率是多少”两部分共同决定。

前者由组合数 C_n^x 决定

$$C_n^x = \frac{n!}{x!(n-x)!} \quad (1)$$

其中， n 为正整数； x 为非负整数， $x = 0, 1, 2, \dots, n$ 。本书前文提过 $!$ 是阶乘运算符，比如 $n!$ 代表

$$n! = 1 \times 2 \times 3 \times \dots \times (n-1) \times n \quad (2)$$

(1) 告诉我们在 n 次试验里选出 x 次成功的排列方式有多少种。



排列与组合的具体含义、公式和计算方式，请大家参考本书附录。

后者则来自伯努利试验本身：成功一次的概率是 p ，失败一次的概率是 $1 - p$ ，因此任意一种“恰好成功 x 次、失败 $n - x$ 次”的具体序列出现的概率为

$$p^x (1-p)^{n-x} \quad (3)$$

将这两部分相乘，我们便得到二项分布的概率质量函数 PMF：

$$p_X(x) = C_n^x p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n \quad (4)$$

这个公式不仅简洁优美，也清晰展示了概率随参数变化的规律。组合数部分体现了“有多少种可能的路径”，而后面的概率乘积部分体现了“每条路径本身发生的可能性”。

图 3 所示为 p 取值为 $1/2$ 、 n 取值为 30 的二项分布 PDF。

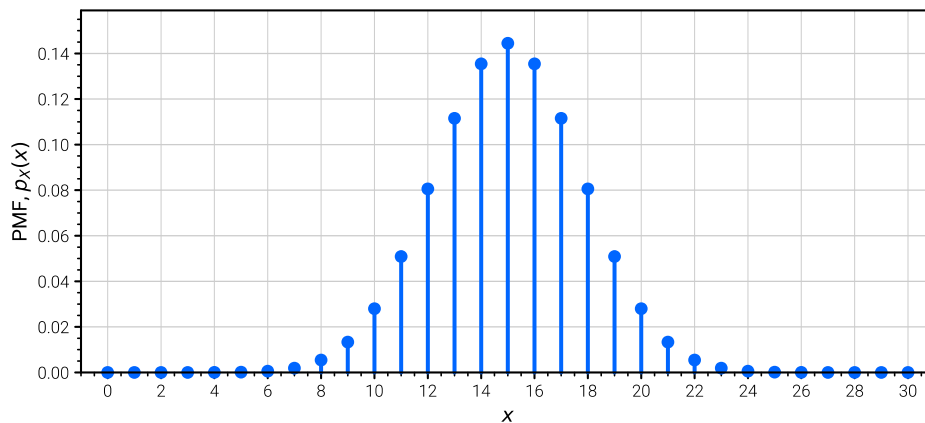


图 3. 二项分布 PMF, $p = 1/2$, $n = 30$



PS_02_06_01.ipynb 提供图 3 的可视化代码，请大家自行学习。

当二项分布的参数取 $n = 1$ 时，它自然就“退化”为伯努利分布，这是因为二项分布的概率质量函数在这种特殊情况下只描述一次试验的成功或失败。

二项分布的所有这些概率加起来恰好为 1，

$$\begin{aligned} \sum_x p_X(x) &= C_n^0 p^0 (1-p)^n + C_n^1 p^1 (1-p)^{n-1} + \dots + C_n^n p^n (1-p)^0 \\ &= (p + (1-p))^n = 1 \end{aligned} \quad (5)$$

这说明它完整地描述了“成功次数”可能出现的全部情况，是一个满足归一化的合法概率分布。

如果 X 服从 (4) 中给出的二项分布， X 的期望和方差分别为：

$$\begin{aligned} E(X) &= n \cdot p \\ \text{var}(X) &= n \cdot p(1-p) \end{aligned} \quad (6)$$

p 对概率质量函数的影响

图 4 展示了当试验次数固定为 $n = 10$ 时，二项分布的概率质量函数随成功概率 p 不同而发生的整体形状变化。可以看到，分布的峰值（即最可能出现的成功次数）以及分布的偏斜方向，都随着 p 的变化而明显改变。

当 p 较大（如 0.8 或 0.9）时，成功更容易发生，因此概率质量集中在较大的 x 附近，图形整体向右偏。

随着 p 逐渐降低到 0.5 时，分布变得相对对称，峰值落在 $x = 5$ 。

当 p 更低（如 0.2 或 0.1）时，成功变得不易发生，概率质量集中在较小的成功次数，图形整体向左偏。

整体对比图 4 这些子图，可以直观感受到：二项分布的形状由参数 p 决定， p 越大，分布越靠右； p 越小，分布越靠左； $p = 0.5$ 时最接近对称。

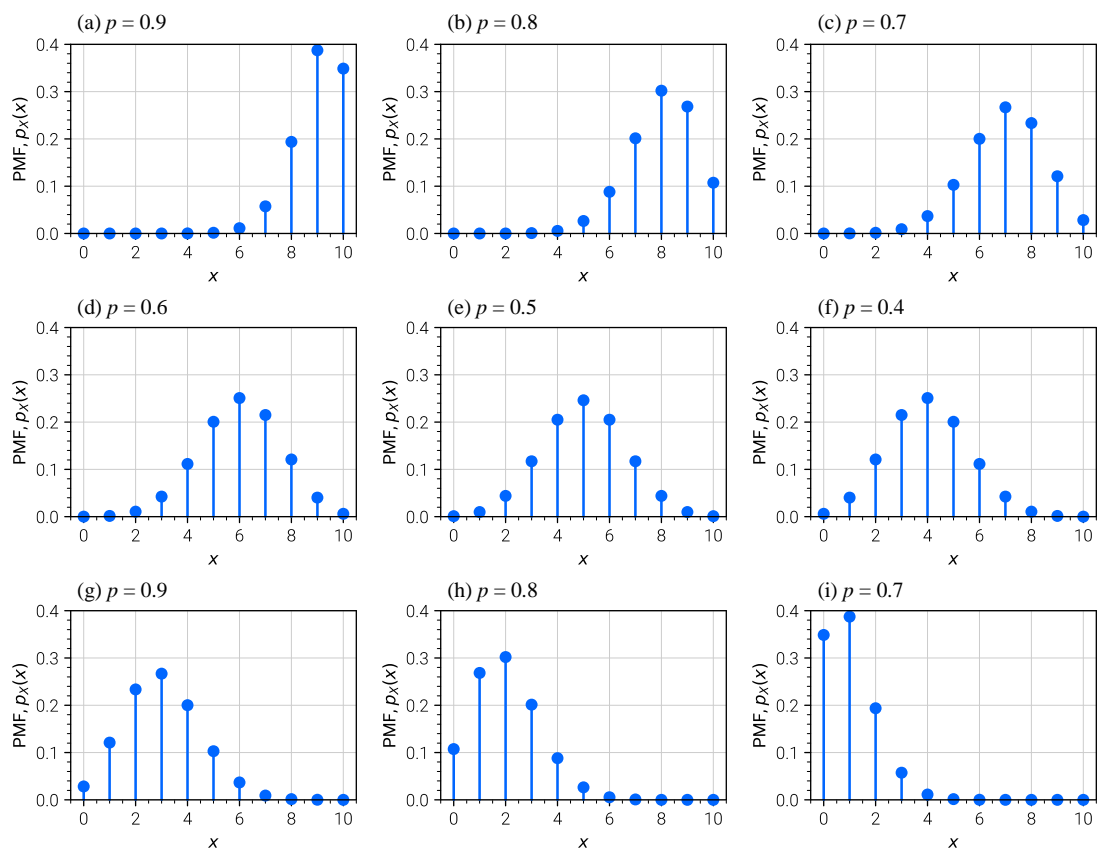


图 4. 二项分布 PMF, p 取不同值, $n = 10$

累积分布函数

图 5 展示了在试验次数同样固定为 $n = 10$ 时，二项分布的累积分布函数 CDF 随成功概率 p 的变化情况。

由于 CDF 描述的是“成功次数不超过 x 的累计概率”，每一条曲线都会从 0 逐级上升到 1，但上升的速度和阶梯出现的位置会因 p 的不同而表现出明显差异。

当 p 较大 (如 0.8 或 0.9) 时, 成功次数更可能集中在较大的区间, 因此 CDF 在前半段上升缓慢, 在接近较大 x 的位置才突然快速跃升到 1; 这说明“少量成功”的概率非常小, 而“成功次数很大”的累计概率增长十分迅速。

随着 p 降低到 0.5 时, 阶梯的上升位置大致集中在中间区域, 体现出分布的对称性和居中趋势。

当 p 更小 (如 0.1 或 0.2) 时, CDF 在较小的 x 就快速上升, 意味着出现较少成功次数的概率非常高, 而对更大 x 的累计概率增加得很慢。

从整体来看, CDF 的阶梯形态清楚地反映了成功概率 p 对分布中心位置的影响: p 越大, 阶梯越偏向右侧; p 越小, 阶梯越偏向左侧; $p = 0.5$ 时阶梯大致位于中间, 呈现旋转对称性。

图 5 的 CDF 与图 4 的 PMF 图互为补充: PMF 告诉我们概率“落在哪儿”, 而 CDF 告诉我们概率“累积到哪儿”。两者结合能够更直观地理解二项分布随参数 p 变化的整体行为。

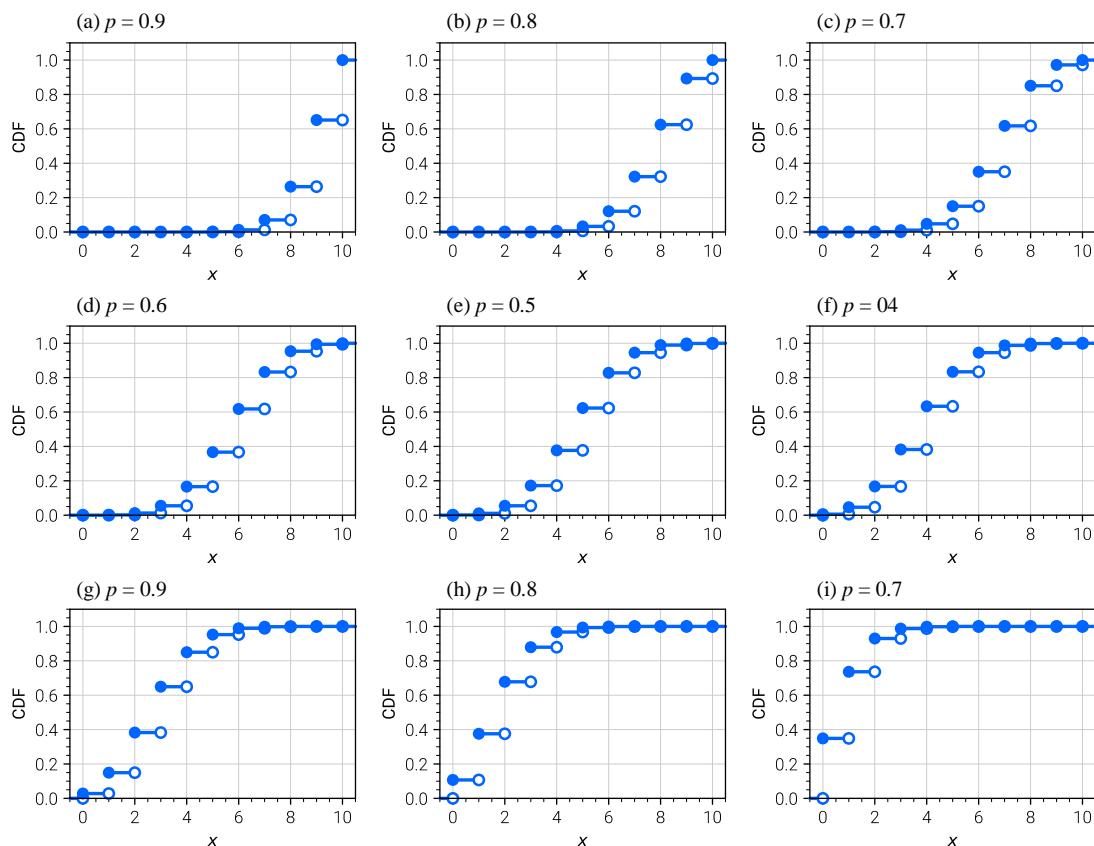


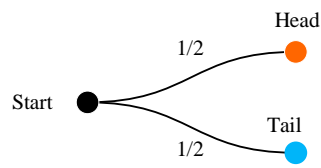
图 5. 二项分布 CDF, p 取不同值, $n = 10$

质地均匀硬币

为了方便大家进一步理解二项分布, 让我们回到本节开篇讲过的抛硬币试验。

为了使问题尽可能直观, 我们先假设硬币是“质地均匀的”, 也就是说正面朝上的概率为 $p = 0.5$ 。如果只抛一次硬币, 这个试验就只有两种可能的结果: 正面或反面。

此时我们令随机变量 X 表示“正面朝上的次数”, 显然 X 只能取 0 或 1。图 6 中展示的两条路径, 正是一次抛硬币可能遇到的全部情况。

图 6. 一枚硬币抛 1 次，2 条路径， $p = 0.5$

对应的概率质量函数非常简单：

$$p_X(x) = \begin{cases} 1/2 & x = 0 \\ 1/2 & x = 1 \end{cases} \quad (7)$$

$X = 1$ 的概率是 0.5， $X = 0$ 的概率也是 0.5。这其实就是最基本的伯努利分布。

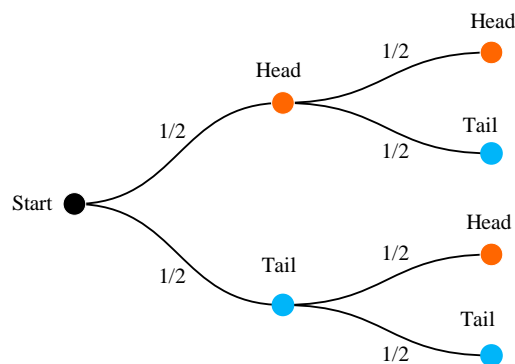
如果一次随机试验中把硬币抛两次，如图 7 所示，可能的路径从两条变成了四条，因为每次抛硬币都独立地有两种结果。随机变量 X 现在表示“两次试验中出现正面的次数”，于是 X 可以取 0、1、2 三个值。

对应的概率可以直接从路径数来理解：例如出现 1 次正面意味着出现“正反”或“反正”，一共两条路径，所以概率是 $2/4 = 0.5$ 。同样地，0 次正面和 2 次正面各只有一条路径，对应的概率是 $1/4$ 。

这样， X 的概率质量函数为：

$$p_X(x) = \begin{cases} 1/4 & x = 0 \\ 1/2 & x = 1 \\ 1/4 & x = 2 \end{cases} \quad (8)$$

这就是二项分布最直观的来源——通过“数路径”得到的概率。

图 7. 一枚硬币抛 2 次，4 条路径， $p = 0.5$

当抛三次硬币以及更多次时，情况也完全类似。

每一次抛硬币都提供两个选择，因此抛 n 次便会产生 2^n 条路径。我们仍然令随机变量 X 表示“正面朝上的次数”。无论 n 是多少，某个具体的取值 $X = x$ 所对应的概率，等于所有“刚好出现 x 次正面”路径的数量除以总路径数。

出现 x 次正面意味着：在 n 个抛硬币的位置中，有 x 个位置取正面，其余取反面，而这些位置的选择方式正好有 C_n^x 种，于是总概率便写成经典的二项分布形式：

$$p_X(x) = \begin{cases} C_n^0 \cdot (1/2)^n & x=0 \\ C_n^1 \cdot (1/2)^n & x=1 \\ \dots & \dots \\ C_n^n \cdot (1/2)^n & x=n \end{cases} \quad (9)$$

图 8 展示了当 $p = 0.5$ 时，不同 n 对应的概率质量函数形状。

随着 n 增加，你会发现图像越来越“鼓起”，形状也越来越接近我们熟悉的高斯分布。需要特别强调的是，二项分布描述的是离散随机变量，图中是概率质量函数；而高斯分布对应连续随机变量，它的函数是概率密度函数，两者不能混淆。

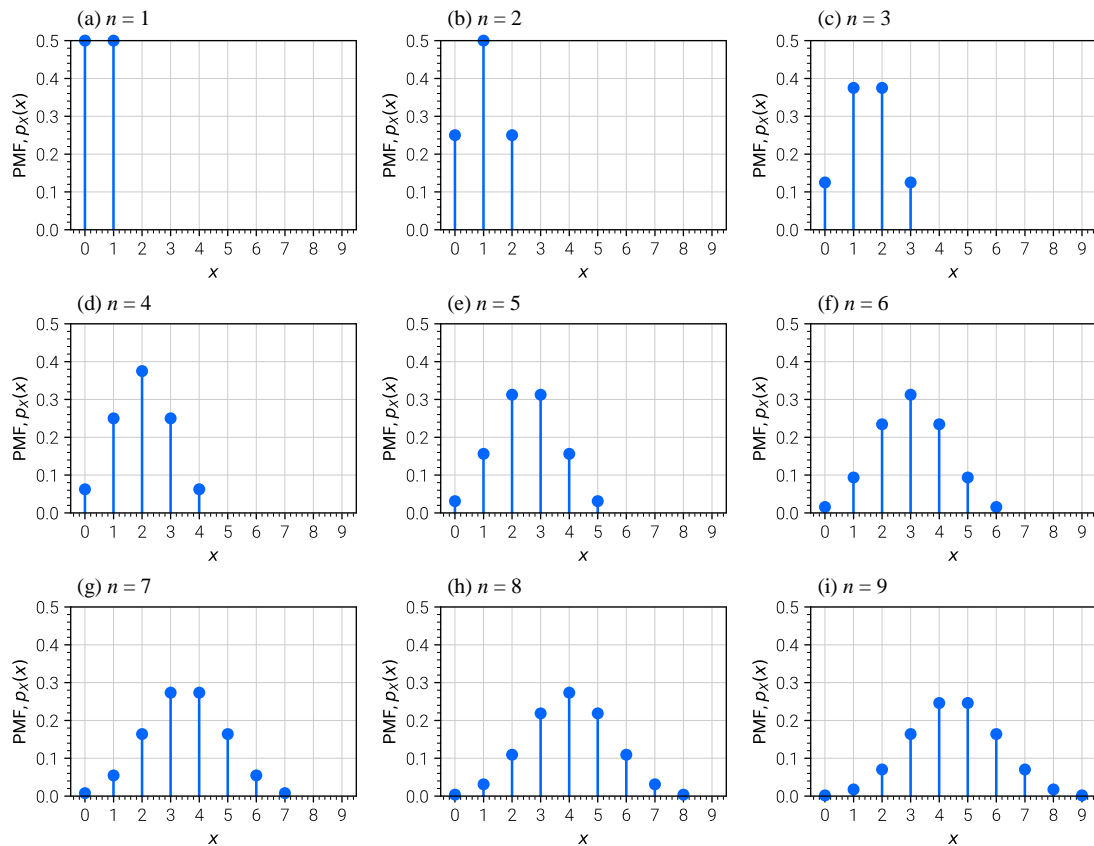
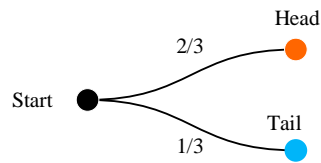


图 8. 二项分布 PMF, n 取不同值, $p = 0.5$

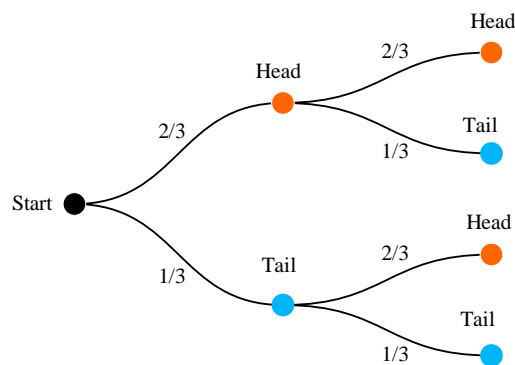
质地不均匀硬币

接下来我们将硬币设为“质地不均匀的”，正面概率不再是 0.5，而是 $p = 2/3$ 。这种硬币更容易落在正面，路径图像仍然与之前一样，只不过每条路径的概率不再相同。

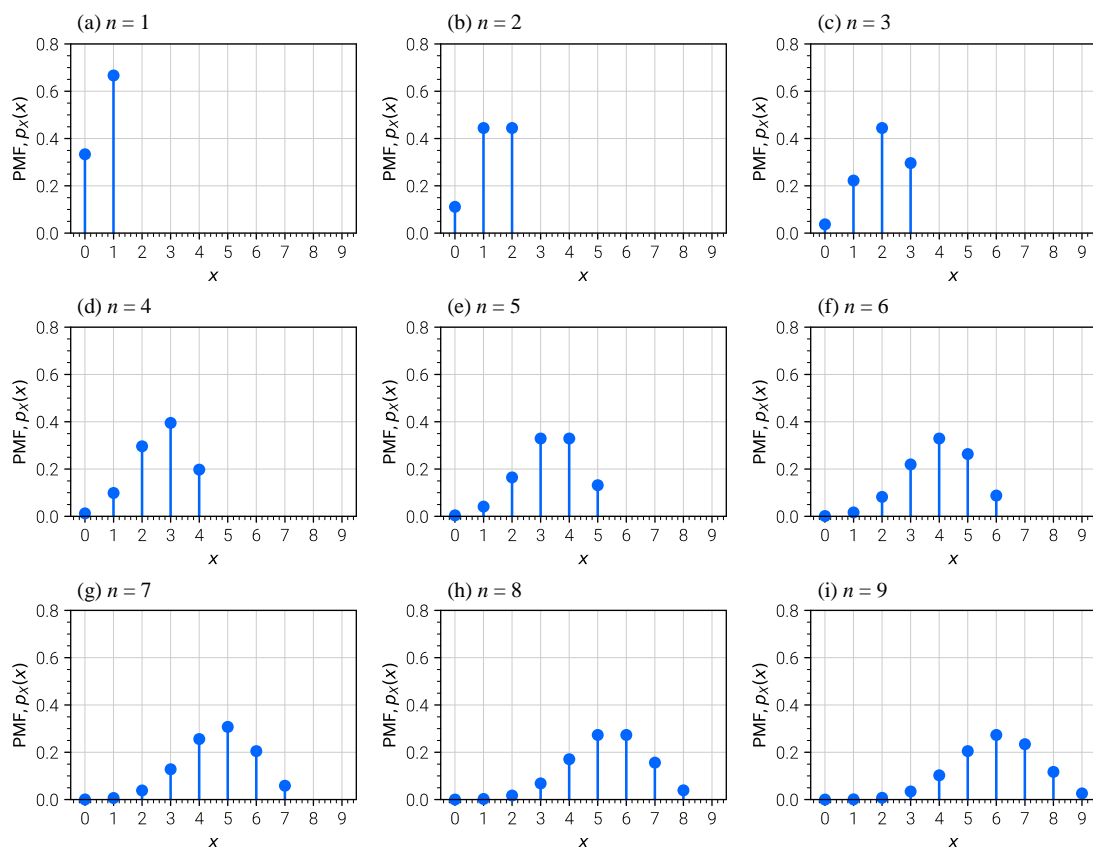
例如抛一次硬币时，正面的概率是 $2/3$ ，反面的概率则是 $1/3$ ，因此这个情形下的随机变量 X 依旧只取 0 或 1，但其概率质量函数与伯努利分布对应的参数 p 已经改变。图 9 展示了对应的两条路径。

图 9. 一枚硬币抛 1 次, 2 条路径, $p = 2/3$

抛两次硬币时, 如图 10 所示, 四条路径的概率不再平均, 而是由每条路径上的“正面/反面组合”共同决定。例如路径“正正”的概率是 $(2/3)^2$, 路径“反反”的概率是 $(1/3)^2$, 而路径“正反”和“反正”的概率都是 $(2/3)(1/3)$ 。统计“出现几次正面”的方式与均匀硬币完全一致: 出现 x 次正面的路径数仍是 C_n^x , 只不过每条路径的概率变成 $p^k(1-p)^{n-k}$ 。

图 10. 一枚硬币抛 2 次, 4 条路径, $p = 2/3$

用相同的方法, 我们可以得到抛三次乃至抛 n 次时的全部概率。图 11 给出了在 $p = 2/3$ 情形下, 不同 n 对应的概率质量函数。当 $p \neq 0.5$ 时, 图像会向出现概率更高的一侧倾斜, 例如当 $p > 1/2$ 时, 图像偏向较大的 x ; 而当 $p < 1/2$ 时, 图像则倾向于 $x = 0$ 附近。

图 11. 二项分布 PMF, n 取不同值, $p = 2/3$ 

请大家用 DeepSeek/ChatGPT 等工具完成本节如下习题。

Q1. 请大家回顾二项式定理，并找到它和二项分布的联系。

Q2. 什么是多项分布，和二项分布有什么联系？

Q3. 什么是几何分布、超几何分布，它们和二项分布有什么联系和区别？