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

2.5 伯努利分布

本节你将掌握的核心技能：

- 离散是“逐点给概率”，连续是“通过面积算概率”；
- 伯努利分布仅有 0 (失败) 和 1 (成功) 两种结果；
- 参数 p 完全决定伯努利分布结构；
- 参数 p 越接近 0.5，分布方差最大，随机性最大；
- 离散随机分布 CDF 跳跃，跳跃高度对应事件概率。

概率分布

在本书前文中，我们已经学习过事件概率，它描述的是在一次随机试验中某个特定结果出现的可能性。然而，仅仅知道事件的概率往往不够。例如，当我们研究一个随机变量时，我们并不只是关心某一个结果发生的概率，而是希望了解它在整个样本空间中各种可能取值的整体分布情况。为了处理这种更加系统化的问题，我们需要引入**概率分布** (probability distribution) 这一概念。

概率分布本质上是一种函数，它量化一个随机变量在不同数值上的“可能性“。简单来说，概率分布告诉我们随机变量可能落在哪里，以及各个区域出现的可能性大小。

和数学中用 $y = ax^2 + bx + c$ 来刻画一条抛物线的几何形状类似，概率分布也是一种经过抽象和理想化的数学模型，用来刻画随机现象背后的规律。

本书前文提过，不同的随机变量其取值方式可能截然不同。离散随机变量只能取有限个或可数个分离的点，连续随机变量则会在某个连续区间内“滑动”。因此，根据随机变量的性质，我们通常把概率分布分为两大类：离散分布 (图 1 (a)、(c))、连续分布 (图 1 (b)、(d))。

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
本书配套微课视频均发布在 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

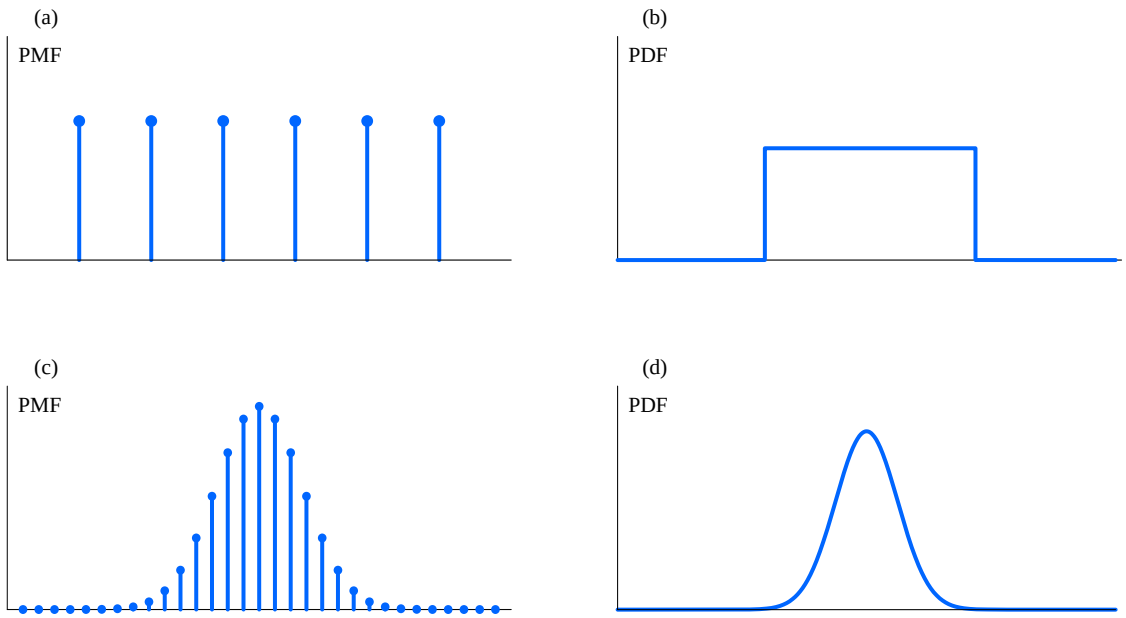

图 1. 对比概率质量函数、概率密度函数

当随机变量的取值是一系列单独的、可以逐个列举的数值时，我们称之为离散分布。典型例子包括抛硬币、掷骰子、统计顾客到达人数等。在这类问题中，概率分布由**概率质量函数** (Probability Mass Function, PMF) 定义，它为每一个可能的取值分配一个概率，如图 1 (a)、(c) 所示。因为离散随机变量的结构是“一点一个概率”，因此计算某个范围内事件发生的概率，只需要把对应点的概率加起来即可。

与之相对，如果随机变量的取值是在实数轴上的某段连续区间，例如温度、身高或机器运行时间，那么我们面对的是连续分布。连续随机变量的概率分布无法像离散情况那样直接为每个点赋予一个概率，具体如图 1 (b)、(d) 所示。因此我们使用**概率密度函数** (Probability Density Function, PDF) 来描述其分布规律。概率密度可理解为“单位长度区间上的概率强度”，某个区间的概率通过对密度函数进行积分获得。

从直观角度看，PMF 更像是一根根竖线 (火柴梗图)，给每个离散点一个确定的高度；而 PDF 则通过曲线下方面积的大小来表达概率。

本节聊聊**伯努利分布** (Bernoulli distribution)，一种常用离散分布。

概率质量函数

理解伯努利分布，就好像理解一个“最小版本”的随机模型，它只有两个可能的结果，却为许多更复杂的概率模型奠定了基础。

伯努利分布描述的是这样一种情形：一次随机试验中只有两种互斥结果，例如硬币要么是正面要么是反面，射击要么命中要么未命中。

为了进行数学建模，我们通常用一个离散随机变量 X 来表示试验的结果，并约定将某个结果记为 1 (例如“成功”或“正面”)，将另一个结果记为 0 (例如“失败”或“反面”)。

这样，整个试验的概率结构就由一个参数 p 决定，它表示事件 $X = 1$ 发生的概率；而事件 $X = 0$ 的概率自然是 $1 - p$。

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
本书配套微课视频均发布在 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

伯努利的概率质量函数可以写成：

$$p_X(x)=\begin{cases} p & x=1 \\ 1-p & x=0 \end{cases} \tag{1}$$

其中，p 满足 $0<p<1$。图 2 所示为伯努利分布的概率质量函数图像。

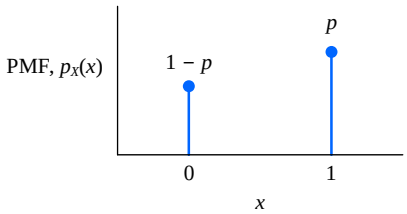

图 2. 伯努利分布的概率质量函数

和所有离散分布一样，伯努利分布的概率质量函数满足归一化条件：

$$\sum_x p_X(x)=p+(1-p)=1 \tag{2}$$

这是概率分布最基本的要求，表示该随机变量的所有可能性已经被完整刻画。

进一步，伯努利分布的概率质量函数还可以写成

$$p_X(x)=p^x(1-p)^{1-x} \quad x\in\{0,1\} \tag{3}$$

其中 x 只能取 0 或 1，参数 p 满足 $0<p<1$；这样保证概率取值非负，且不大于 1。

如果将 $x=1$ 代入上式，可以得到 $\Pr(X=1)=p$；将 $x=0$ 代入，则得到 $\Pr(X=0)=1-p$。这种写法的好处是，通过一个统一的表达式同时涵盖两种可能结果。

注意，(3) 这种形式会在本册后续贝叶斯推断中出现。

p 对概率质量函数的影响

图 3 用火柴梗图展示了伯努利分布在不同参数 p 取值下的概率质量函数随 p 的变化情况。

图 3 中共有九个子图，分别对应 $p=0.9, 0.8, 0.7, \ldots, 0.1$。每个子图的横轴是随机变量 x，取值只能是 0 或 1；纵轴表示 PMF，即伯努利分布在这两个点上的概率。

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
本书配套微课视频均发布在 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

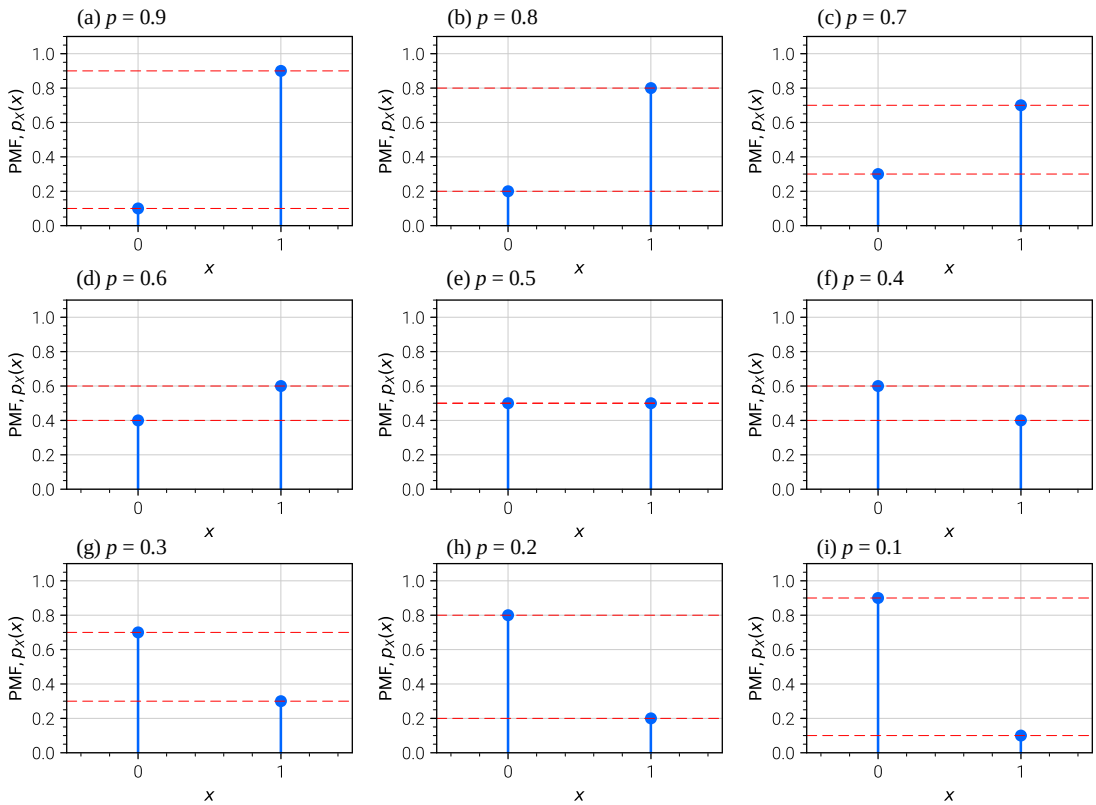

图 3. 伯努利分布 PMF，p 取不同值

PS_02_05_01.ipynb 提供图 3 的可视化代码，请大家自行学习。

期望

伯努利分布的期望为

$$\mathrm{E}(X)=p \tag{4}$$

让我们简单推导一下。

前文提过，离散随机变量 X 的期望为

$$\mathrm{E}(X)=\sum_{x} x\cdot p_X(x) \tag{5}$$

带入伯努利的概率分布，得到

$$\mathrm{E}(X)=0\times(1-p)+1\times p=p \tag{6}$$

图 4 所示为 p 取不同值时，伯努利分布的期望值位置。

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
本书配套微课视频均发布在 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

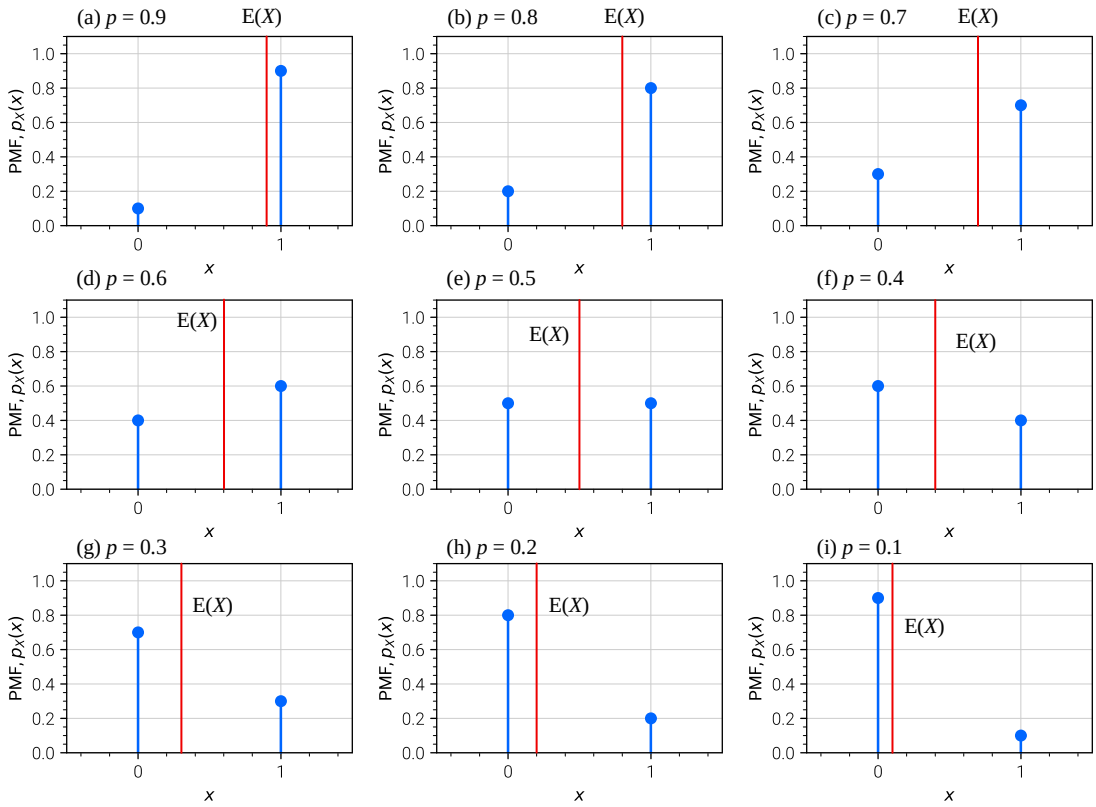

图 4. 伯努利分布 PMF 的期望值位置，p 取不同值

期望值可以理解为“大量重复试验后 1 出现的平均比例”。

本书前文介绍的抛一枚硬币的试验就是常见的伯努利分布。如果硬币质地均匀，获得正面 ($X = 1$)、反面 ($X = 0$) 的概率均为 0.5，则 X 的概率质量函数为：

$$p_X(x) = \begin{cases} 0.5 & x = 1 \\ 0.5 & x = 0 \end{cases} \tag{7}$$

容易计算得到期望为 0.5。

图 5 展示了在大量重复抛掷硬币的过程中，样本均值如何随着试验次数的增加逐渐趋近于真实概率。这幅图对应的是一枚均匀硬币，正反面出现的概率各为 0.5。由于随机性的影响，样本均值在最开始波动较大，但随着抛掷次数增长，这些波动迅速被“平均掉”，最终曲线缓慢稳定地逼近 0.5。

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
本书配套微课视频均发布在 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

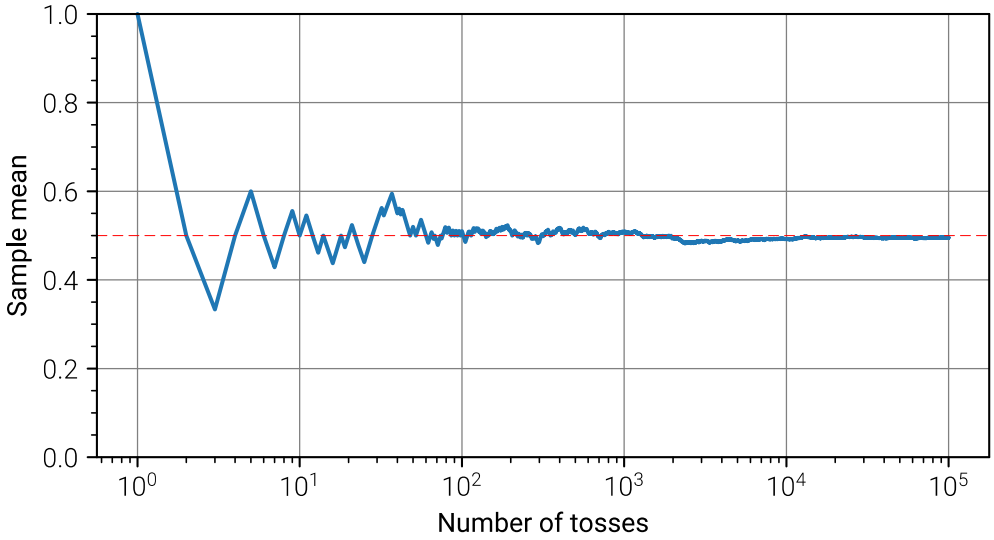

图 5. 抛质地均匀硬币 10000 次，均值随次数变化

如果硬币质地不均匀，头重脚轻；假设获得正面的概率为 0.6，则对应获得背面的概率为 1 – 0.6 = 0.4。则 X 的概率质量函数为：

$$p_X(x)=\begin{cases}0.6 & x=1\\ 0.4 & x=0\end{cases} \tag{8}$$

图 6 对应的则是一枚质地不均匀硬币，正面出现的概率为 0.6。整体趋势与均匀硬币的情形类似：在试验初期，均值随结果的随机变化而剧烈起伏，但随着抛掷次数增加，样本均值逐渐收敛，并稳定地贴近真实概率 0.6。

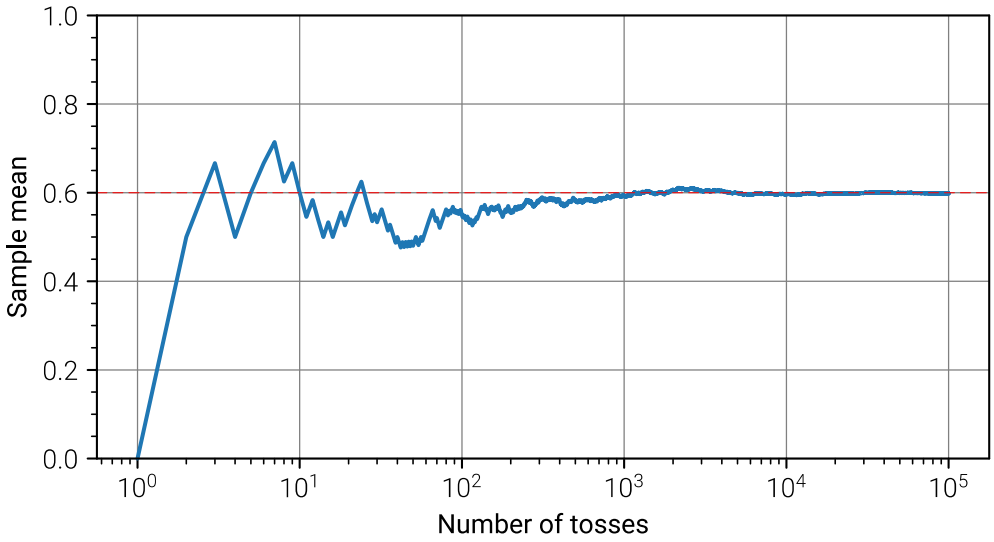

图 6. 抛头重脚轻硬币 10000 次，均值随次数变化

方差

伯努利分布的方差为

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
本书配套微课视频均发布在 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

$$\mathrm{var}(X) = p(1-p) \tag{9}$$

本书前文提过，对于离散随机变量 X，其方差可以通过下式计算得到

$$\mathrm{var}(X) = \underbrace{\mathrm{E}(X^2)}_{\text{Expectaton of } X^2} - \underbrace{\mathrm{E}(X)^2}_{\text{Square of } \mathrm{E}(X)} \tag{10}$$

让我们先计算 X^2 的期望

$$\mathrm{E}(X^2) = 0^2 \times (1-p) + 1^2 \times p = p \tag{11}$$

带入 (10) 得到

$$\mathrm{var}(X) = p - p^2 = p(1-p) \tag{12}$$

方差用来衡量结果的不确定性。如图 7 所示，当 p 越接近 0.5 时，随机性最强，因此方差最大；当 p 越接近 0 或 1 时，结果趋于确定，方差也随之变小。

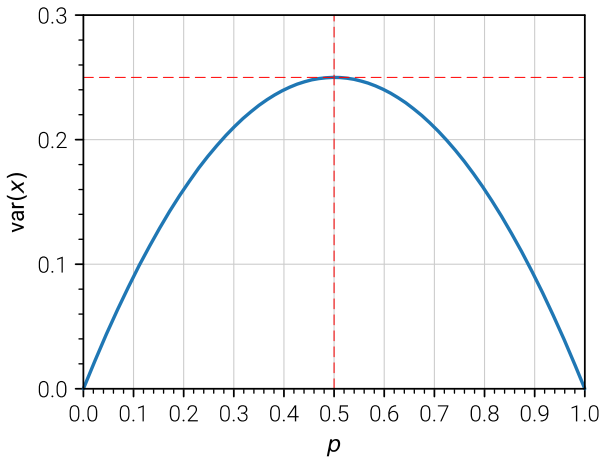

图 7. 伯努利分布方差随 p 变化

累积分布函数

本章前文提过离散随机变量 X 的累积分布函数 $F_X(x)$ 定义为：

$$F_X(x) = \Pr(X \le x) = \sum_{t \le x} p_X(t) \tag{13}$$

因此，伯努利分布的 CDF 可以写成：

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1-p & 0 \le x < 1 \\ 1 & 1 \end{cases} \tag{14}$$

通过解析式，我们发现这个分段函数有两个跳跃点：在 $x = 0$ 时从 0 跳跃到 $1-p$；在 $x = 1$，从 $1-p$ 跳跃到 1，这个跳跃的大小正好是事件成功概率 p。

图 8 展示了伯努利分布的 CDF 随不同的成功概率 (p) 的变化情况。每个子图对应一个不同的 p 值，从 0.9 到 0.1 依次递减。

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
本书配套微课视频均发布在 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

在每个子图中，当 $x < 0$ 时，CDF 始终为 0，因此图中左侧是沿着 0 水平线的一段。

在 $x = 0$ 处，CDF 会跃升到 $1 - p$，图中以空心圆表示“右极限” (不包含点)。

在 $0 \leqslant x < 1$ 区间，CDF 保持不变，始终等于 $1 - p$，也就是图中的水平蓝线。

在 $x = 1$ 处，CDF 再次发生跳跃，从 $1 - p$ 跃升到 1 (跳跃幅度为 p)，图中的实心蓝点表示这个跳跃终点。

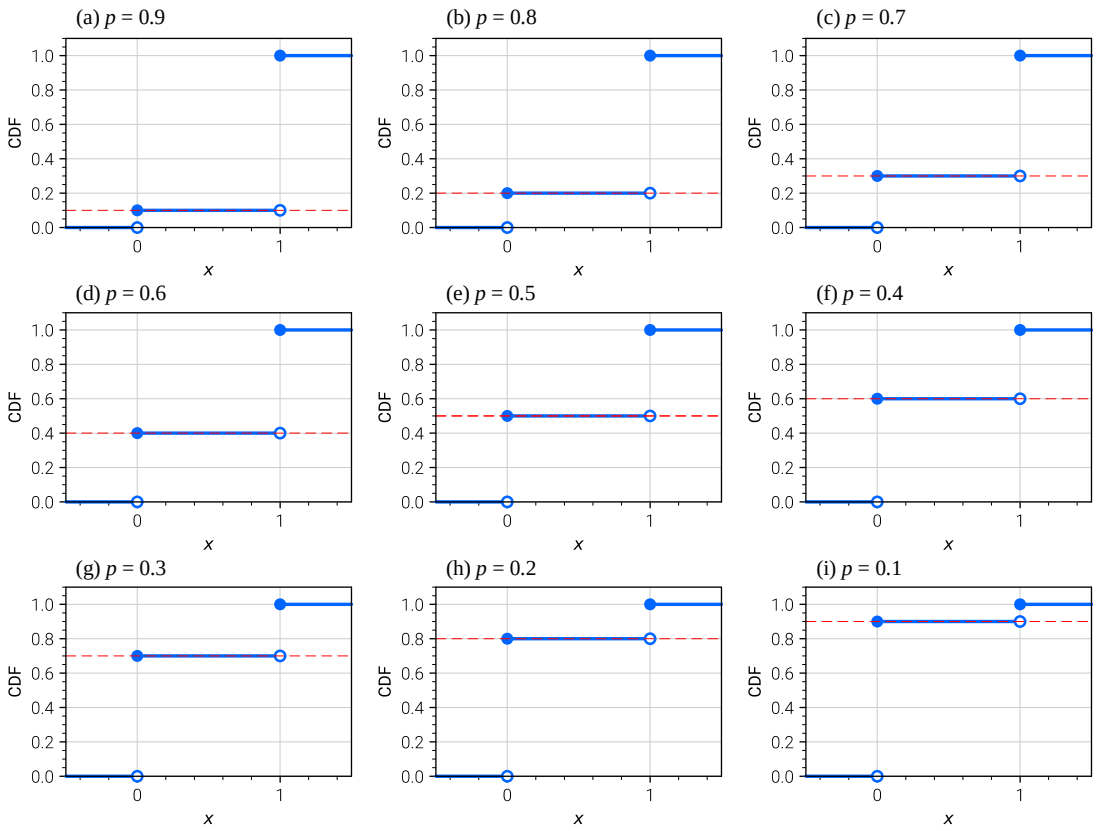

图 8. 伯努利分布 CDF，p 取不同值

观察离散随机变量的 CDF 时可以发现，函数在每个可能取值处都会突然上升一段，这段上升的高度正好等于该随机变量在该点的概率质量。

因此，CDF 在这些点上的跳跃大小就暗含了全部的概率信息，而在两点之间的区间内，CDF 始终保持不变。这意味着，理解这些断点的位置及其对应的跳跃幅度，就等同于完全理解了这个离散分布。

请大家用 DeepSeek/ChatGPT 等工具完成本节如下习题。

Q1. 若一枚硬币正面概率为 0.4，求伯努利随机变量的期望和方差。

Q2. 比较 $p = 0.1$ 与 $p = 0.9$ 的伯努利分布，讨论哪种情况下分布的方差更大。

Q3. 用 Python 写代码模拟抛质地均匀硬币，重复抛掷 1000 次，并统计 1 出现的频率。

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
本书配套微课视频均发布在 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