

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

## 2.7 泊松分布



### 本节你将掌握的核心技能：

- ▶ 泊松分布描述单位时间或空间内事件发生次数的离散概率分布；
- ▶ 事件之间相互独立，且在观察区间内事件发生的平均速率保持恒定；
- ▶ 泊松分布的期望和方差都等于  $\lambda$ ；
- ▶  $\lambda$  越小，泊松分布越陡峭且右偏； $\lambda$  越大，分布越趋于对称。

### 泊松分布

泊松分布 (Poisson distribution) 是一类非常重要的离散概率分布，用来刻画事件在单位时间或单位空间内发生的次数。

这里的“事件”通常具有两个特点：一是不同事件之间相互独立，不会互相影响；二是在观察区间内，事件发生的平均速率保持不变。

事件发生的平均速率通常记为  $\lambda$ ，它表示在单位时间或空间中，事件预期会出现多少次。

为了帮助理解，可以先想象一下常见场景：比如一小时内呼叫中心接到的电话数量，一天里某个路口发生的交通事故次数，或一个网页每秒收到的访问请求数。它们都是不确定的随机事件，但只要整体规律保持稳定，平均发生速度  $\lambda$  就可以很好地描述它们的行为。

泊松分布的概率质量函数为：

$$p_x(x) = \frac{\exp(-\lambda)\lambda^x}{x!}, \quad x = 0, 1, 2, \dots \quad (1)$$

其中， $x$  是事件发生的次数，为非负整数 ( $x = 0, 1, 2, 3, 4, 5, \dots$ )； $\lambda$  是单位时间内事件平均发生的次数。

**⚠ 注意**， $\lambda$  为正实数，也就是说  $\lambda$  未必是正整数，可以是 (正) 小数，比如 5.8。

图 1 所示为  $\lambda$  为 4 时泊松分布的 PMF 火柴梗图。再次强调,  $x$  可以取得任意非负整数; 当  $x$  较大时概率质量函数趋近于 0 (但是不等于 0), 为了方便可视化图 1 中  $x$  取值限定在 0 到 20 之间。

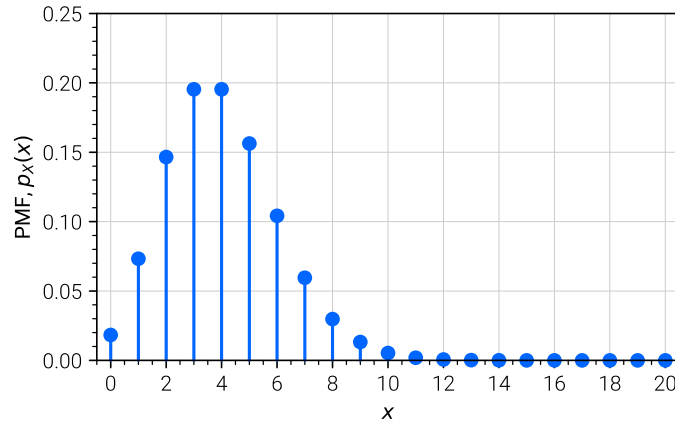


图 1. 泊松分布 PMF,  $\lambda = 4$



PS\_02\_07\_01.ipynb 提供图 1 的可视化代码, 请大家自行学习。

泊松分布的概率质量函数刻画的是: “在给定的平均速率  $\lambda$  下, 一段时间或空间里恰好发生  $x$  次事件的概率是多少”。

满足 (1) 泊松随机变量的期望和方差都是  $\lambda$ :

$$E(X) = \text{var}(X) = \lambda \quad (2)$$

换句话说, 泊松分布的平均行为和其波动程度完全由同一个参数决定。这意味着平均发生率越高, 其波动性 (不确定性) 也越大。

## 举个例子

某公交站点平均每小时会有 4 辆公交车经过 ( $\lambda = 4$ )。下一小时恰好有 0 辆公交车经过的概率约为:

$$\Pr(X = 0) = \frac{e^{-4} 4^0}{0!} = \frac{1}{1} e^{-4} \approx 0.0183 \quad (3)$$

下一小时恰好到 5 辆车的概率为:

$$\Pr(X = 5) = \frac{e^{-4} 4^5}{5!} = \frac{1024}{120} e^{-4} \approx 8.533 \times 0.0183 \approx 0.156 \quad (4)$$



请大家自行计算当  $x = 1, 2, 3, 4, 6, 7, 8$  等值时的概率值。

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>

本书配套微课程视频均发布在 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](mailto:jiang.visualize.ml@gmail.com)

### $\lambda$ 对概率质量函数的影响

图 2 展示了泊松分布随参数  $\lambda$  变化的概率质量函数 PMF 特征。泊松分布 PMF 峰值总是接近  $\lambda$ ，这是因为泊松分布期望值和  $\lambda$  相同。随着  $\lambda$  增大，峰值右移。泊松分布 PMF 方差也等于  $\lambda$ ，因此  $\lambda$  越大，分布越宽，概率集中区越平缓，对称性增强。

在图 2 (a) 中  $\lambda = 1$ ，泊松分布 PMF 大部分概率集中在  $x = 0$  和  $x = 1$ ，极大概率在低值，峰值陡峭，表现为事件偶尔发生。

随着  $\lambda$  增大，如在图 2 (b)  $\lambda = 2$ 、(c)  $\lambda = 3$ ，分布峰值向右移动，概率集中区逐渐展开，表示事件更频繁发生。

图 2 (e)-(i) 对应  $\lambda = 5$  到 9，分布趋向于更宽和对称，峰值继续右移，概率在较大的  $x$  上也有分布，长尾逐渐缩短，泊松分布逐渐近似于正态分布。当  $\lambda$  较小时，分布强烈右偏 (长右尾)，随着  $\lambda$  增大，右偏减少，渐趋对称。

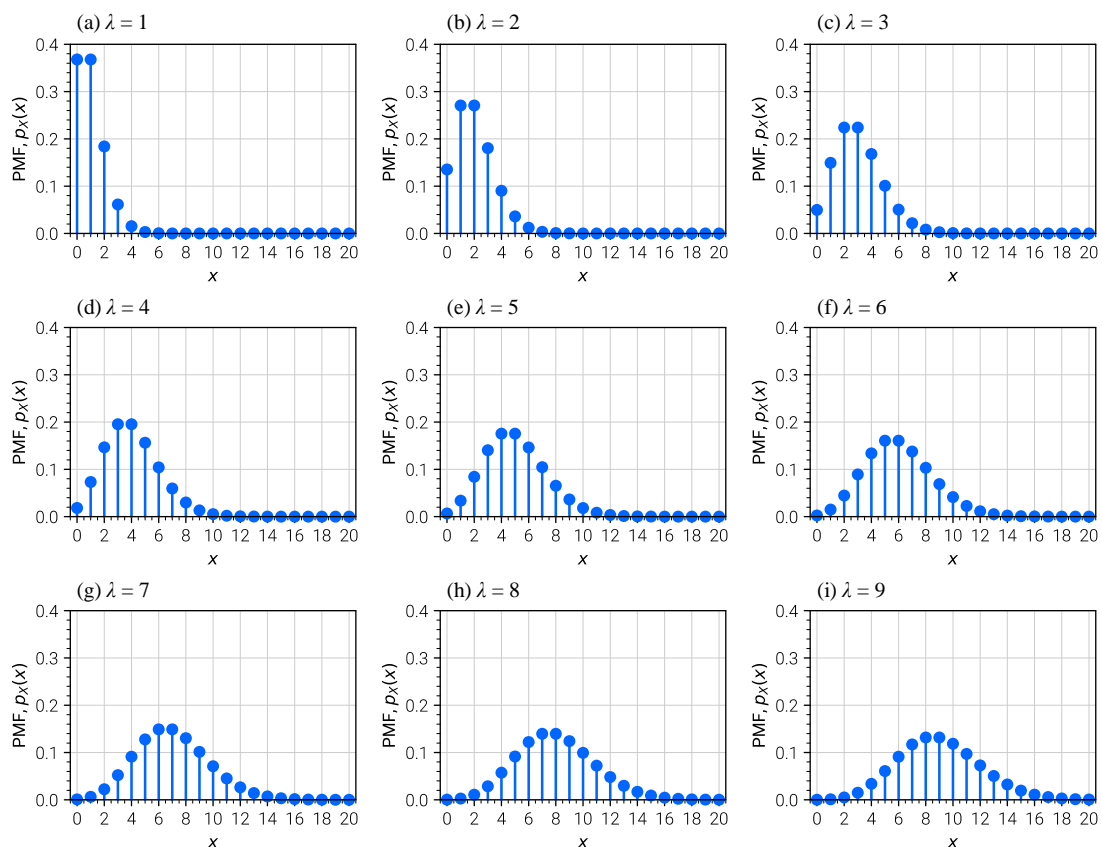


图 2. 泊松分布 PMF,  $\lambda$  取不同值

### 累积分布函数

图 3 展示了泊松分布在不同参数  $\lambda$  下的累积分布函数 CDF 形状和增长特征。 $\lambda$  较小时，CDF 增长极快，几乎全部概率在极少数点上累积； $\lambda$  较大时，CDF 增长变慢、分布区间变宽。

图 3 (a) 中  $\lambda = 1$ ，CDF 在  $x = 0, 1, 2$  处迅速上升，几乎全部概率集中在较小的  $x$ ，曲线很快趋于 1，说明事件很少多次发生。

随着  $\lambda$  从 4 递增至 9，CDF 曲线进一步右移，S 型越平缓，需要更大的  $x$  值来累计到接近 1，表明随机变量具有更大的期望和方差。

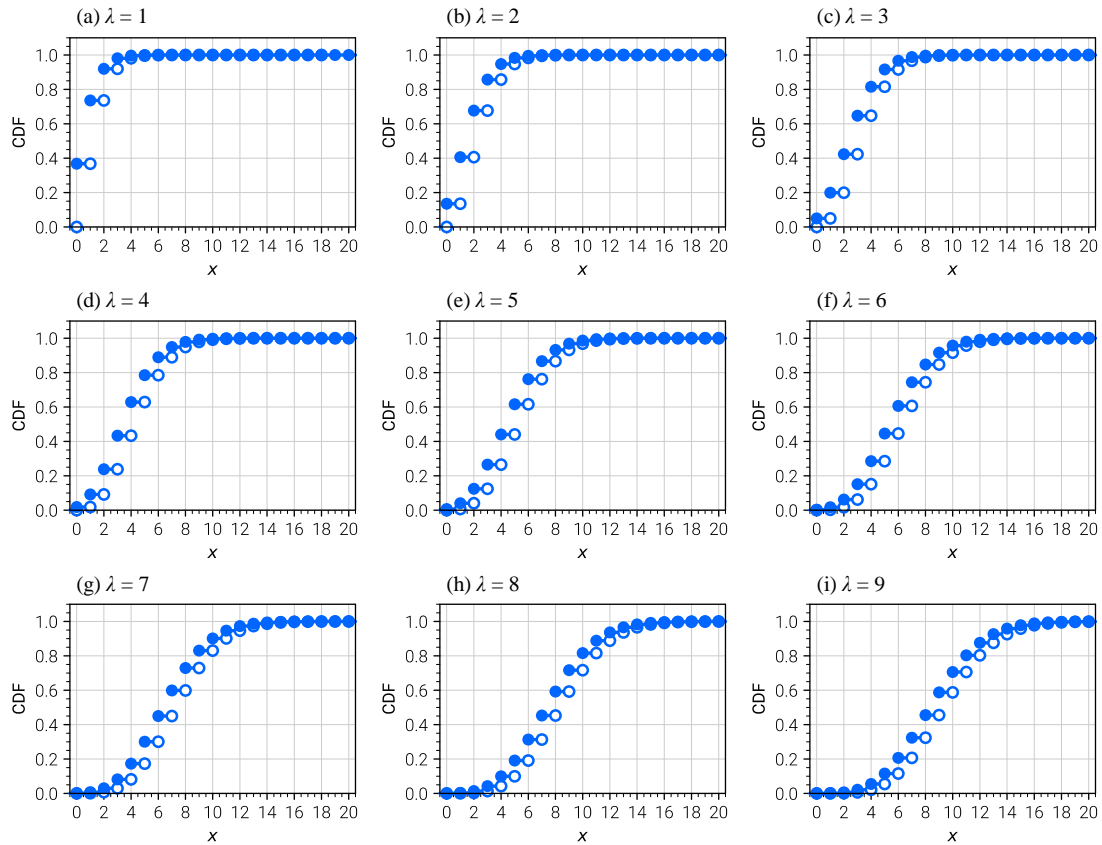


图 3. 泊松分布 CDF,  $\lambda$  取不同值



请大家用 DeepSeek/ChatGPT 等工具完成本节如下习题。

**Q1.** 用 Python 绘制  $\lambda = 2$ 、 $\lambda = 5$ 、 $\lambda = 8$  三个泊松分布的概率质量函数对比图， $x$  轴范围设为 0 到 15，用不同颜色区分，并添加图例。

**Q2.** 某呼叫中心平均每分钟接到 2.5 通电话。计算：

- ▶ 某一分钟恰好接到 3 通电话的概率；
- ▶ 某一分钟接到不超过 3 通电话的概率；
- ▶ 某一分钟接到至少 4 通电话的概率。

**Q3.** 用 `numpy.random.poisson()` 生成  $\lambda = 6$  的泊松分布随机样本 10000 个，学习如何计算样本的均值和方差，并与理论值 ( $\lambda = 6$ ) 进行比较。

**Q4.** 当  $n$  趋于无穷大时，如何从二项分布推导得到泊松分布？