

概率论与数理统计-基础实践

0. 简介

报告提交截止时间：2021 年 11 月 6 日

报告要求：格式请参考模板，篇幅 4 页左右

编程语言：推荐使用 Python，但不作限制

使用 Python 语言完成该作业时可能会用到的库：

(1) numpy, 包含大量科学计算相关的函数, 在数据分析和机器学习中十分常用, 详情请参考:

<https://numpy.org/>

(2) scipy, 其中的 scipy.stats 包含多种概率与统计学相关的函数, 详情请参考:

<https://docs.scipy.org/doc/scipy/reference/stats.html>

(3) matplotlib, 其中 matplotlib.pyplot 是 Python 中最常用的 2D 画图工具, 详情请参考:

https://matplotlib.org/3.1.1/api/pyplot_summary.html

(4) mpl_toolkits, mpl_toolkits.mplot3d.Axes3D 包含绘制 3D 图形相关的工具, 详情请参考:

<https://matplotlib.org/stable/tutorials/toolkits/mplot3d.html>

1. 掷骰子实验

问题：用程序模拟掷骰子的 n 次独立重复实验, 记录每次骰子朝上的点数, 统计各个点数出现的频率。分别模拟当 $n = 10$, $n = 1,000$ 和 $n = 100,000$ 时, 各个点数朝上的频率。

说明与提示：实验结果可以用终端 print 显示, 也可以用绘制图像的方式呈现 (样图如图 1 所示)。Python 语言中的 matplotlib.pyplot 库包含绘制各种平面图形的函数, 具体使用方法请参考:

https://matplotlib.org/3.1.1/api/pyplot_summary.html

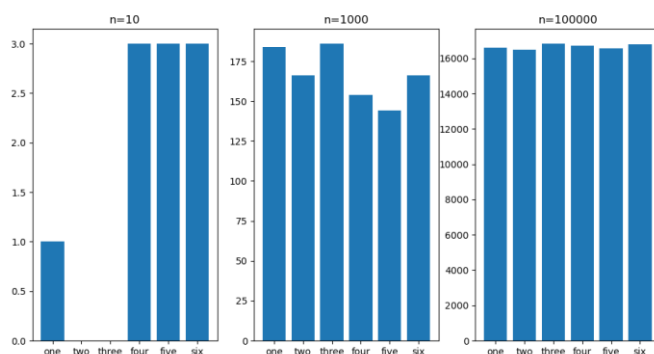


图 1. 掷骰子实验结果频数统计

2. 概率密度与分布函数

问题：泊松定理: 设 $\lambda > 0$ 是一个常数, n 是任意正整数, 设 $np_n = \lambda$, 则对任意一个固定的非负整数 k , 有

$$\lim_{n \rightarrow \infty} \binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{\lambda^k e^{-\lambda}}{k!}.$$

泊松定理表明, 当 n 很大时, 以 n , p 为参数的二项分布的概率值可以由参数为 $\lambda = np$ 的泊松分布的概率值近似。

假设 λ 固定为8, 请绘制出当 (n, p) 分别为 $(10, 0.8)$, $(40, 0.2)$, $(160, 0.05)$ 时的二项分布的分布律图, 并分别绘制出可以逼近这些二项分布的泊松分布的分布律图 (样图如图 2 所示)。观察当 n 的大小变化时, 该泊松分布的拟合效果如何变化。

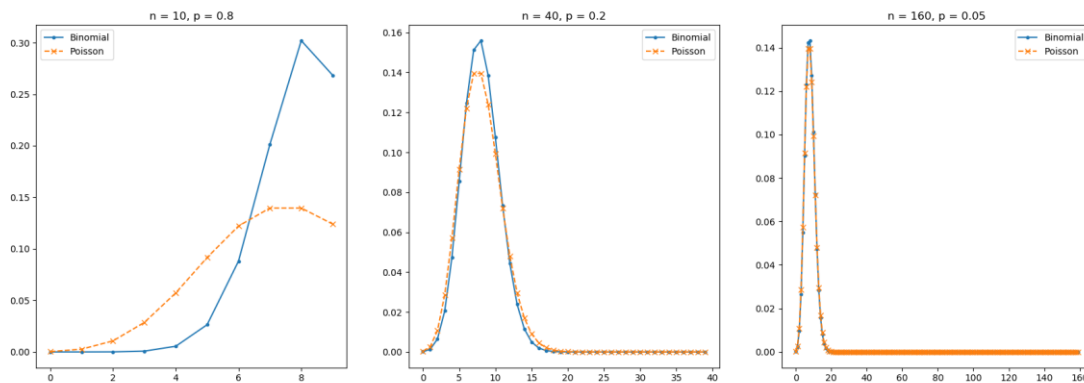


图 2. 用泊松分布近似二项分布

说明与提示：该题既可以自己实现两种分布函数的概率密度，也可以调用 Python 中的 `scipy.stats` 库。`scipy.stats` 库包含各种分布函数可以直接使用，十分方便，具体使用方法请参考：

<https://docs.scipy.org/doc/scipy/reference/stats.html#>

3. 二维正态分布及其边缘分布

问题：二维正态分布 $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ 的概率密度为

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ \frac{-1}{2(1-\rho^2)} \left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right] \right\}$$

其中 $\sigma_1 > 0$, $\sigma_2 > 0$, $-1 < \rho < 1$ 。对于二维正态随机变量 (X, Y) ，当且仅当 $\rho = 0$ 时， X 和 Y 互相独立。该二维正态分布中，无论 ρ 的取值如何变化，关于 X 的边缘分布为 $X \sim N(\mu_1, \sigma_1^2)$ ，关于 Y 的边缘分布为 $Y \sim N(\mu_2, \sigma_2^2)$ 。

现假设 $\mu_1 = 0.3$, $\mu_2 = 0.5$, $\sigma_1 = 0.3$, $\sigma_2 = 0.5$ 。

(1) 当 $\rho = 0$ 时，以 3D 方式绘制出二维正态分布图形，并分别绘制出关于 X 的边缘分布和关于 Y 的边缘分布（样图如图 3 所示）；

(2) 当 $\rho = 0.9$ 和 $\rho = -0.9$ 时，旋转 3D 图形到合适的视角，观察二维正态分布的中心与关于 X 的边缘分布中心和关于 Y 的边缘分布中心之间的关系。

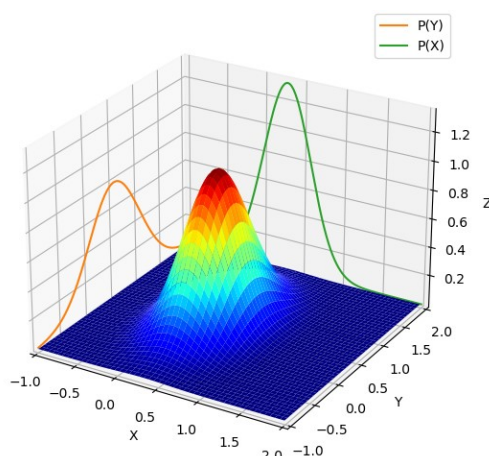


图 3. 二维正态分布及其边缘分布

说明与提示：Python 语言中的 `scipy.stats` 库中包含多维正态分布的函数

`scipy.stats.multivariate_normal(mean=None, cov=1)`

对于二维正态分布而言，参数中的 $\text{mean} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ ， $\text{cov} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$ ，详细使用方法请参考：

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.multivariate_normal.html#scipy.stats.multivariate_normal

Python 语言中的 `mpl_toolkits.mplot3d` 库中包含绘制 3D 图形的函数，详细使用方法请参考：

<https://matplotlib.org/stable/tutorials/toolkits/mplot3d.html#d-plots-in-3d>