

二项分布极限近似的Python可视化仿真及教学应用

伊克热木·买买提依明

新疆大学数学与系统科学学院, 新疆 乌鲁木齐

收稿日期: 2026年2月27日; 录用日期: 2026年3月27日; 发布日期: 2026年4月3日

摘要

二项分布的泊松近似和正态近似, 是概率论与数理统计课的核心内容, 能帮助学生更好地理解中心极限定理。本文用Python的scipy.stats和matplotlib库, 做了二项分布与泊松、正态分布的叠加对比图, 还实现了误差定量评估和动态收敛演示。通过典型参数的数值实验, 验证了两种近似的有效性, 分析了误差特点及适用条件, 也检查了连续性校正对正态近似精度的改善。本文的可复现实验平台和可视化工具, 有实际教学价值。

关键词

二项分布, 泊松近似, 正态近似, 中心极限定理, 可视化教学, Python

A Visual Simulation and Teaching Application Study of the Limit Approximation of the Binomial Distribution Based on Python

Ikram Mamtimin

School of Mathematics and Systems Science, Xinjiang University, Urumqi Xinjiang

Received: February 27, 2026; accepted: March 27, 2026; published: April 3, 2026

Abstract

The Poisson and normal approximations to the binomial distribution are key topics in probability and mathematical statistics courses, helping students grasp the central limit theorem better. This paper uses Python's scipy.stats and matplotlib libraries to create overlaid comparison charts between the

binomial distribution and its Poisson or normal counterparts, along with quantitative error assessment and dynamic convergence demos. Through numerical experiments with typical parameters, it verifies the effectiveness of both approximations, analyzes their error features and suitable conditions, and checks how continuity correction improves the normal approximation's accuracy. The reproducible experiment platform and visualization tools here offer real value for teaching.

Keywords

Binomial Distribution, Poisson Approximation, Normal Approximation, Central Limit Theorem, Visualization in Teaching, Python

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

二项分布作为刻画独立伯努利试验中成功次数的经典概率模型，在概率论与数理统计课程体系中具有基础性地位。其在特定参数极限条件下分别可由泊松分布与正态分布进行近似，这不仅降低了大样本情形下的计算复杂度，也为理解分布收敛与中心极限定理提供了典型范式[1][2]。然而，在传统教学实践中，相关内容多停留于公式推导与静态结论，学习者往往难以直观把握“由离散到近连续”的收敛机制及不同参数区间下的近似误差特征[3]。

近年来，统计教育逐步转向计算支撑与可视化驱动的教学范式[4]-[6]。已有研究指出，基于图形交互与动态演示的教学方式能够有效提升学生对概率分布极限行为与近似条件的理解深度[7]。基于此，本文依托 Python 科学计算生态构建可复现的数值实验框架，结合分布叠加可视化、误差指标量化与动态收敛展示，系统呈现二项分布向泊松分布与正态分布的近似路径与精度变化规律，从而为概率统计课程教学提供可迁移的实践方案。

2. 离散型与连续型随机变量

离散型随机变量的概率规律用概率质量函数(PMF)描述，连续型随机变量的分布特征用概率密度函数(PDF)刻画，二者之间的渐近联系是理解分布近似和中心极限定理的关键路径。本节通过 Python 可视化方法展示典型分布的形态，为后面的近似分析提供图形知觉和比较基础。

为直观比较离散型分布与连续型分布，本文选取了参数相近的一组进行可视化：二项分布取 $n=20$, $p=0.3$ (均值 $np=6$, 方差 $np(1-p)=4.2$)；泊松分布取 $\lambda=6$ ；正态分布取 $N(6, 4.2)$ 。用 Python 的 `scipy.stats` 和 `matplotlib` 算出各分布的 PMF/PDF，并叠加在同一坐标系(见图 1)。这个设计为后续误差分析和近似判断提供了直观依据。

- (1) 二项分布($n=20, p=0.3$): 呈轻度右偏的离散单峰形态；
- (2) 泊松分布($\lambda=6$): 同为离散单峰，偏态程度与二项分布接近；
- (3) 正态分布($N(6, 4.2)$): 连续、对称的钟形曲线。

从图 1 可以看出，在中等均值条件下，二项分布和泊松分布在峰位与整体轮廓上很相似，正态分布则是更平滑的连续特征。这个对比显示了离散分布向连续近似过渡的路径。教学中，可以通过动态调节参数(如增大 n 或 λ)来观察分布形态演化，从而加深对极限收敛过程的理解。

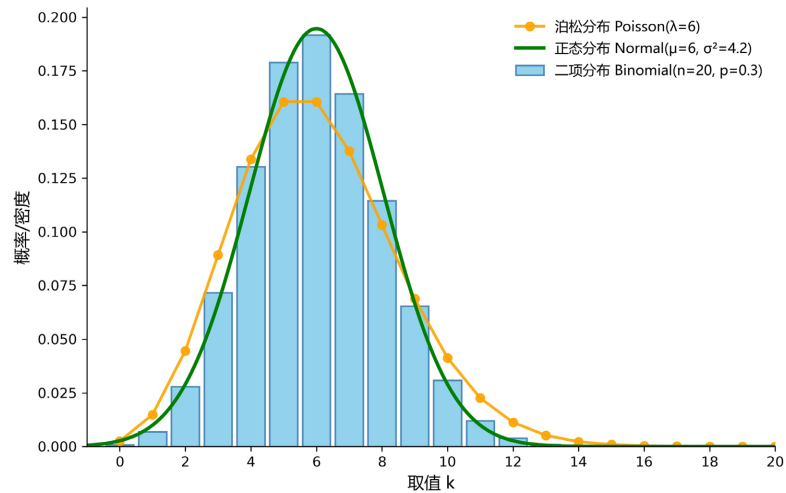


Figure 1. Visual comparison of typical probability distributions
图 1. 典型概率分布可视化对比

3. 二项分布的泊松与正态近似：数值实现与精度评估

二项分布的概率质量函数为

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, \dots, n \quad (1)$$

期望是 $E(X) = np$ ，方差是 $Var(X) = np(1-p)$ 。其极限行为有“双路径收敛”特征[8][9]： $n \rightarrow \infty, p \rightarrow 0$ 且 $\lambda = np$ 有限时收敛到泊松分布 $Poisson(\lambda)$ ；当 $n \rightarrow \infty$ 且 $p \in (0, 1)$ 固定时收敛到正态分布 $N(np, np(1-p))$ 。

二项分布在不同参数极限下的收敛路径，是离散分布向连续分布过渡的关键，其主要结论见表 1。

Table 1. Limit convergence path of the binomial distribution

表 1. 二项分布的极限收敛路径

极限路径	质量	排序
二项→泊松	$n \rightarrow \infty, p \rightarrow 0, np = \lambda$ (固定)	泊松 $P(\lambda)$
二项→正态	$n \rightarrow \infty, p \in (0, 1)$ 固定	正态 $N(np, \sqrt{np(1-p)})$
泊松→正态	$\lambda \rightarrow \infty$	正态 $N(\lambda, \lambda)$

其结构化示意图 2。

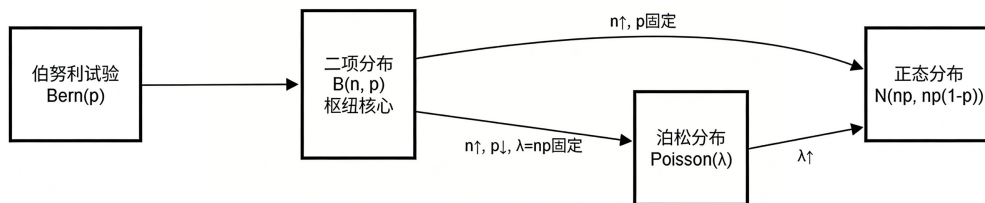


Figure 2. The limiting convergence path of the binomial distribution

图 2. 二项分布的极限收敛路径

本文以二项分布为核心，搭建了一个教学框架，通过理论推导、数值模拟和可视化展示，来系统比较两条近似路径的收敛机制、误差特征和适用参数区间。

3.1. 泊松近似：理论基础、实现与精度分析

3.1.1. 理论基础

当 $n \rightarrow \infty$ 、 $p \rightarrow 0$ 且 $\lambda = np$ 为有限时，二项分布 $\text{Bin}(n, p)$ 收敛到泊松分布 $\text{Poisson}(\lambda)$ [1] [2]，其概率质量函数是

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, k = 0, 1, 2, \dots \quad (2)$$

这个结果为低概率、高重复次数的计数建模提供了理论基础和计算便利，所以在多种稀有事件计数问题中有应用价值[10]。

3.1.2. 数值示例与精度评估

选择参数 $n = 100$ 、 $p = 0.05$ ($\lambda = 5$)。表 2 列出了 $k = 0$ 到 15 的主要概率值和绝对偏差。

Table 2. Accuracy of the Poisson approximation ($n = 100$, $p = 0.05$, $\lambda = 5$)

表 2. 泊松近似精度 ($n = 100$, $p = 0.05$, $\lambda = 5$)

k	二项分布概率	泊松分布概率	绝对偏差	k	二项分布概率	泊松分布概率	绝对偏差
0	0.00592	0.00674	0.00082	8	0.06518	0.06527	0.00009
1	0.03116	0.03369	0.00253	9	0.03690	0.03626	0.00064
2	0.08118	0.08422	0.00304	10	0.01907	0.01813	0.00094
3	0.13958	0.14037	0.00009	11	0.00901	0.00824	0.00077
4	0.17814	0.17546	0.00074	12	0.00391	0.00343	0.00048
5	0.18094	0.17546	0.00548	13	0.00156	0.00132	0.00024
6	0.15037	0.14622	0.00415	14	0.00057	0.00047	0.00010
7	0.10592	0.10444	0.00148	15	0.00019	0.00016	0.00003

从表 2 看，大多数 k 点的绝对偏差小于 0.005，最大偏差是 0.00548。在 $k = 0 \sim 15$ 的截断区间，总变差距离约 0.013，这表明泊松近似在该参数下精度较高。

3.1.3. 可视化分析

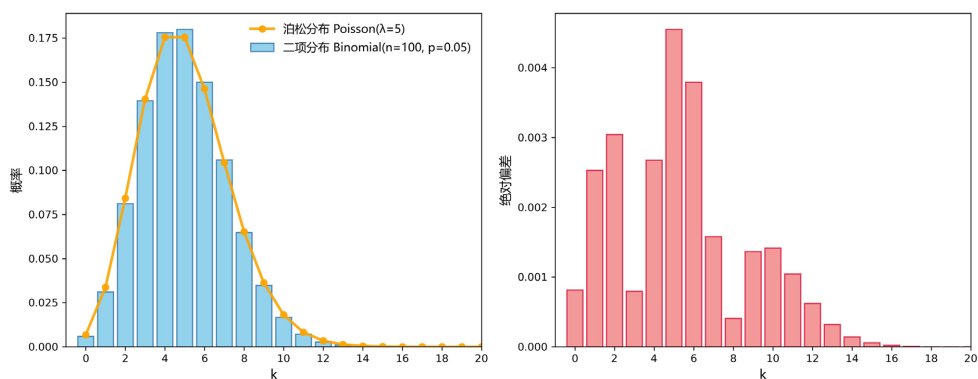


Figure 3. Comparative analysis of the binomial and Poisson distributions: approximation and error analysis ($n = 100$, $p = 0.05$, $\lambda = 5$)

图 3. 二项分布与泊松分布的近似对比及偏差分析 ($n = 100$, $p = 0.05$, $\lambda = 5$)

图 3 给出了叠加对比和误差分布。左图可以看到两者整体重合度较高；右图显示绝对偏差在 λ 附近达到峰值，随后快速衰减。

3.1.4. 精度影响因素与适用条件

泊松近似的精度同时受 n 和 p 影响。在本文的实验参数范围内， n 较大且 p 较小时，泊松近似通常能满足教学所需精度。随着 n 增大，误差呈下降趋势，这与理论预期一致。

3.2. 正态近似：理论基础、实现与精度分析

3.2.1. 理论基础

当 $n \rightarrow \infty$ 时，二项随机变量标准化后收敛到标准正态分布，即

$$P\left(\frac{X - np}{\sqrt{np(1-p)}} \leq x\right) \rightarrow \Phi(x), \tag{3}$$

其中 $\Phi(x)$ 为标准正态分布函数[1]。

在有限样本下可采用连续性修正[1]：

$$P(X = k) \approx \Phi\left(\frac{k + 0.5 - \mu}{\sigma}\right) - \Phi\left(\frac{k - 0.5 - \mu}{\sigma}\right), \tag{4}$$

其中 $\mu = np$ ， $\sigma = \sqrt{np(1-p)}$ 。该处理可减小离散到连续转换带来的偏差。

3.2.2. 数值示例与精度评估

选取参数 $n = 150$ ， $p = 0.3$ ($\mu = 45$ ， $\sigma \approx 5.612$)。表 3 给出了代表性 k 点的概率和绝对偏差。

所选区间内绝对偏差不超过 0.0012，说明在该参数下，结合连续性修正的正态近似精度较高。

Table 3. Accuracy of the normal approximation ($n = 150$, $p = 0.3$, $\mu = 45$, $\sigma \approx 5.612$)

表 3. 正态近似精度 ($n = 150$, $p = 0.3$, $\mu = 45$, $\sigma \approx 5.612$)

k	二项分布概率	正态近似概率	绝对偏差	k	二项分布概率	正态近似概率	绝对偏差
34	0.01022	0.01045	0.00023	50	0.04674	0.04779	0.00105
36	0.01988	0.01969	0.00019	52	0.03205	0.03268	0.00063
38	0.03346	0.03268	0.00078	54	0.01955	0.01969	0.00014
40	0.04897	0.04779	0.00119	56	0.01063	0.01045	0.00018
42	0.06263	0.06156	0.00107	58	0.00516	0.00489	0.00028
44	0.07026	0.06987	0.00039	60	0.00224	0.00201	0.00023
46	0.06939	0.06987	0.00048	62	0.00087	0.00073	0.00014
48	0.06052	0.06156	0.00104	64	0.00030	0.00023	0.00007

3.2.3. 可视化分析

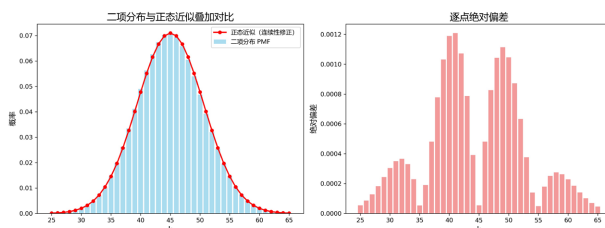


Figure 4. Overlay comparison and error analysis of the binomial distribution and its normal approximation ($n = 150$, $p = 0.3$)

图 4. 二项分布与正态近似叠加对比及偏差分析 ($n = 150$, $p = 0.3$)

图 4 展示了叠加结果和误差分布。加入连续性修正后，主要区间的重合度较高，误差在尾部快速衰减。

3.2.4. 精度影响因素与适用条件

正态近似的精度主要由 n 以及 np 、 $n(1-p)$ 决定。常用经验阈值是 np ， $n(1-p) \geq 5$ ，更稳健可取 ≥ 10 [11]。当 n 较小或 p 接近 0 或 1 时，误差会增大。在本文参数范围内，结合连续性修正的正态近似可以达到较高精度。

3.3. 近似适用条件与比较

两类近似的常用判据和误差特征归纳在表 4。

Table 4. Applicability conditions and error characteristics of the Poisson and normal approximations to the binomial distribution

表 4. 二项分布的泊松近似与正态近似适用条件及误差特征

近似类型	常用适用条件	主要误差来源	量化指标(典型边界案例)
泊松	n 较大, p 较小, $np \leq 10$	方差偏差(p 较大时)	最大绝对偏差 ≈ 0.012
正态	$np \geq 5$, $n(1-p) \geq 5$ (稳健建议 ≥ 10)	离散 - 连续截断与偏度	最大绝对偏差 ≈ 0.018

数值实验表明，在推荐参数区间内，两种近似都能达到教学所需精度；超出该范围时，更适合使用精确二项计算。

4. 教学实施与反思

为验证本文可视化工具在教学中的效果，我们做了一个小型实证实验。对象为某高校概率论课程的 40 名本科生，随机分为两组：对照组(20 人)采用传统板书和静态图讲解二项分布近似；实验组(20 人)使用 Python 工具进行交互演示，包括参数调节、分布叠加和误差分析。实验时长 90 分钟，前后通过小测验(5 道选择题 + 2 道简答题，覆盖近似条件、误差和中心极限定理)及 Likert 问卷(1~5 分，评估收敛机制、精度和参数影响)收集数据，并用 t 检验分析，结果见表 5。

Table 5. Quantitative comparison of empirical teaching experiment results

表 5. 实证教学实验结果量化比较

指标	对照组(传统法)	实验组(可视化工具)	统计显著性
课后测验得分提升幅度	15.2%	28.5%	$P < 0.05$
问卷理解自信度提升	0.7 分	1.2 分	$p < 0.05$

数据表明，可视化方法能显著提升学生对分布收敛的理解，尤其有助于把握抽象概念。

在教学应用中，该工具可用于：(1) 课堂演示：实时调节 n 、 p 观察分布演化；(2) 课后作业：学生复现实验并分析误差；(3) 探究项目：扩展到负二项等分布，培养计算能力。这些方式能更好地连接理论与实践，促进主动学习。

挑战主要包括教师需要一定 Python 熟练度、学生可能忽视公式推导，以及设备条件受限。可采用以下措施：(1) 预置模板并开展培训；(2) 结合推导与可视化进行混合教学；(3) 开发 Web 版工具。这样可提升工具适用性，支持概率教育创新。

5. 结论与教学意义

数值结果表明，在常见适用条件下，两类近似都能达到较高精度，最大绝对偏差通常低于 0.02。对

于小样本或重尾情形，宜优先采用精确计算。

本文通过逐点对比、叠加可视化和误差量化，系统展示了“二项→泊松”与“二项→正态”的收敛过程。配套的 Python 可复现实验框架有助于学生动态观察参数变化对分布形态和误差的影响，从而建立“条件-结论-精度”的整体理解。该框架可作为概率论课程中连接两类极限定理的核心教学模块。

参考文献

- [1] 孙舒婷. 二项分布的正态近似若干条件[J]. 理论数学, 2022, 12(9): 1419-1428.
- [2] 徐玉华, 曾明. 泊松分布性质及应用研究[J]. 长江大学学报(自科版), 2006(4): 475-476.
- [3] 曹俊飞, 曾志红. 厚基础重应用——概率论与数理统计课程教学改革的探索与实践[J]. 广东第二师范学院学报, 2018, 38(3): 98-103.
- [4] 金莹. 《数理统计学》课程教学范式改革研究[J]. 河南科技, 2014(12): 274-275.
- [5] 洪珊珊, 田卫忠. 人工智能驱动的概率统计课程改革[J]. 教育进展, 2025, 15(8): 850-858.
- [6] 徐扬, 李成, 徐辰武, 等. 水利类专业中数理统计方法的教学改革与实践应用——评《常用统计软件的应用》[J]. 灌溉排水学报, 2022, 41(10): 160-160.
- [7] 张水胜, 宇世航, 王晓霞. 基于 MATLAB 的概率论极限理论的教学探讨[J]. 高师理科学刊, 2007, 27(4): 78-79.
- [8] 韩非. 计算二项分布与负二项分布期望、方差的新思路[J]. 新乡学院学报(自然科学报), 2008, 25(2): 9-11.
- [9] 程志明. 二项分布的期望和方差的计算方法综述[J]. 宜春学院学报, 2010, 32(12): 21-22.
- [10] 刘瑞元, 张智霞. 二项分布与泊松分布判别的假设检验[J]. 青海大学学报: 自然科学版, 2008, 26(1): 44-47.
- [11] 舒婷孙. 二项分布的正态近似若干条件[J]. 理论数学, 2022, 12(9): 1419-1428.