计算物理 A——Homework 9

何金铭 PB21020660

1 题目描述

考虑泊松分布、指数分布，并再自设若干个随机分布（它们有相同或不同的 μ 和 σ^2），通过 Monte Carlo 模拟，验证中心极限定理成立（N =2、5、10）。

2 理论分析

2.1 分布的选取

2.1.1 Possion 分布

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

且有 $\mu=\lambda,\sigma^2=\lambda,\sigma=\sqrt{\lambda}$

2.1.2 指数分布

$$P(X=x)=\lambda e^{-\lambda x} \tag{2}$$

且有 $\mu=\dfrac{1}{\lambda},\sigma^2=\dfrac{1}{\lambda^2},\sigma=\dfrac{1}{\lambda}$

满足指数分布的随机变量 $x=-\dfrac{1}{\lambda}\log\xi,\xi\in[0,1]$

2.1.3 随机分布 1

再取一个离散分布的函数，不妨就讨论两点分布

$$\begin{cases} P(X=1)=p \\ P(X=0)=1-p \end{cases} \tag{3}$$

且有 $\mu=p,\sigma^2=p(1-p),\sigma=\sqrt{p(1-p)}$

讨论 $p=\dfrac{1}{2}$

则有 $\mu=\dfrac{1}{2},\sigma=\dfrac{1}{2}$

2.1.4 随机分布 2

再取一个连续分布的函数，不妨就讨论 $[0,a]$ 上的均匀分布。

$$P(X=x)=\frac{1}{a}\quad x\in[0,a] \tag{4}$$

且有 $\mu=\dfrac{a}{2},\sigma^2=\dfrac{a^2}{12},\sigma=\dfrac{a\sqrt{3}}{6}$

若取 $a=1$

则有 $\mu=\dfrac{1}{2},\sigma^2=\dfrac{1}{12},\sigma=\dfrac{\sqrt{3}}{6}$

2.2 中心极限定理

Theorem 1. *Lindeberg–Lévy CLT* $\{X_1, \ldots, X_n, \ldots\ N\}$ 是一系列独立随机变量，且有 $E(X_i) = \mu, Var(X_i) = \sigma^2 < \infty$。则 $Y = \frac{\langle X_N \rangle - \mu}{\frac{\sigma}{\sqrt{N}}} \sim N(0,1)$

由中心极限定理可得：

$$P\left(-\lambda < \frac{\langle X_N \rangle - \mu}{\frac{\sigma}{\sqrt{N}}} < \lambda\right) = \int_{-\lambda}^{\lambda} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\, dt = 1 - \alpha \tag{5}$$

所以可得于置信水平为 $1-\alpha$ 时，偏差 $\sigma_s = |\langle X_N \rangle - \mu|$ 满足：

$$|\langle X_N \rangle - \mu| < \lambda \frac{\sigma}{\sqrt{N}} \tag{6}$$

且有 $\langle X_N \rangle = \frac{\sum_{i=1}^{i=N} X_i}{N}$
以下讨论取 $\lambda = 3, 1-\alpha = 0.99$ 时偏差的情况

3 算法过程

1. 对于 N 次抽样，每次随机选取 N 个点，并求出 $\langle X_N \rangle, \mu, \sigma^2$，即可得统计量 $Y = \frac{\langle X_N \rangle - \mu}{\frac{\sigma}{\sqrt{N}}}$
2. 重复操作 1，直至有 M 个符合要求的 Y。
3. 若于 N 很大的情况下成立，$Y = \frac{\langle X_N \rangle - \mu}{\frac{\sigma}{\sqrt{N}}} \sim N(0,1)$，则可验证中心极限定理的成立。

对于不同的随机分布，在具体的处理时有不同的做法。

3.1 Possion 分布

$\lambda = 1$ 时，$\mu = 1, \sigma = 1, Y = \frac{\langle X_N \rangle - 1}{\frac{1}{\sqrt{N}}}$

3.2 指数分布

$\lambda = 1$ 时，$\mu = 1, \sigma = 1, Y = \frac{\langle X_N \rangle - 1}{\frac{1}{\sqrt{N}}}$

3.3 两点分布

$p = \frac{1}{2}$ 时，$\mu = \frac{1}{2}, \sigma = \frac{1}{2}, Y = \frac{\langle X_N \rangle - 0.5}{\frac{1}{2\sqrt{N}}}$

3.4 均匀分布

$a = 1$ 时，$\mu = \frac{1}{6}, \sigma = \frac{\sqrt{3}}{6}, Y = \frac{\langle X_N \rangle - \frac{1}{6}}{\frac{\sqrt{3}}{6\sqrt{N}}}$

4 程序说明

4.1 主要程序

schrage.c 16807 随机数生成器 (*Schrage* 法) 的代码

clt_possion.py 生成 Possion 分布的抽样以及其中心极限定理检验的代码

clt_exp.py 生成指数分布的抽样以及其中心极限定理检验的代码

clt_2p.py 生成两点分布的抽样以及其中心极限定理检验的代码

clt_unit.py 生成均匀分布的抽样以及其中心极限定理检验的代码

4.2 程序结果

schrage.exe 16807 PRNG

16807.csv 一个种子数为"23942907" 的样本随机数文件

possion.csv Possion 分布的抽样结果

exp.csv 指数分布的抽样结果

2p.csv 两点分布的抽样结果

unit.csv 均匀分布的抽样结果

./out 文件夹路径, 里面存放了各个分布的中心极限定理验证的抽样，其中文件名代表了分布函数，后缀数字为 N 的值

./pic 文件夹路径, 里面存放了各个分布的中心极限定理验证的作图，其中文件名代表了分布函数，中间数字为 N 的值，后缀数字为抽样次数 M 的次数

4.3 其他说明

1. 数据都写于 CSV 文件中
2. 其中 Python 程序用到的库有:

- matplotlib.pyplot：用于作图
- numpy：用于数据处理
- csv：用于读写 CSV 文件

5 结果分析

5.1 Possion 分布

$N = 2$ 时

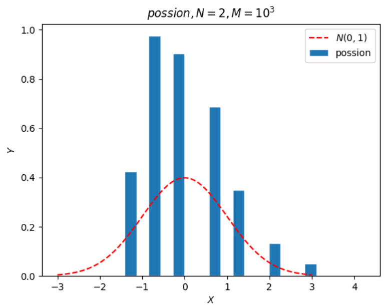

图 1: $M = 10^3$

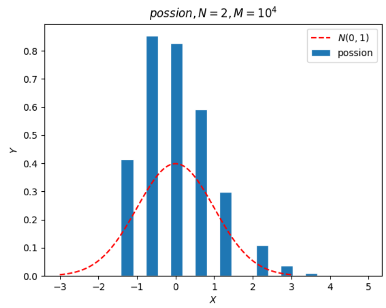

图 2: $M = 10^4$

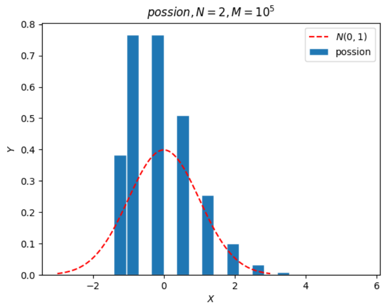

图 3: $M = 10^5$

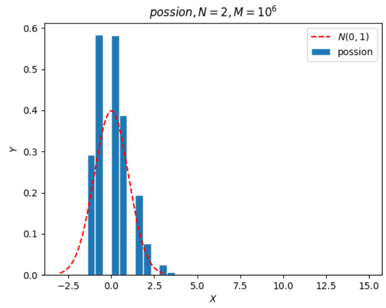

图 4: $M = 10^6$

$N = 5$ 时

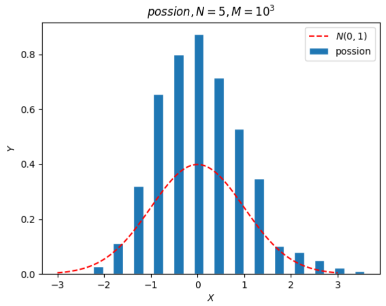

图 5: $M = 10^3$

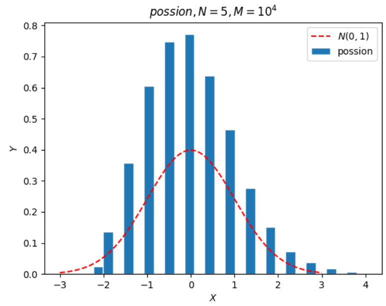

图 6: $M = 10^4$

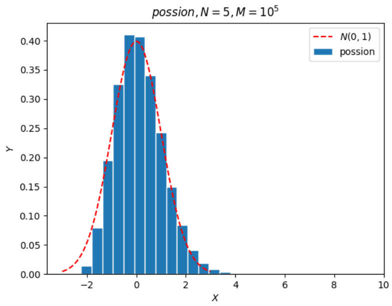

图 7: $M = 10^5$

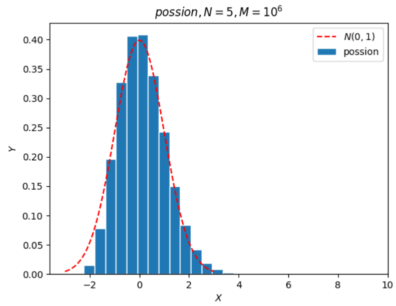

图 8: $M = 10^6$

$N = 10$ 时

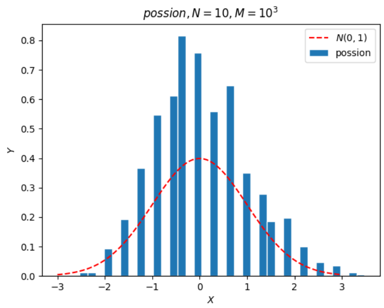

图 9: $M = 10^3$

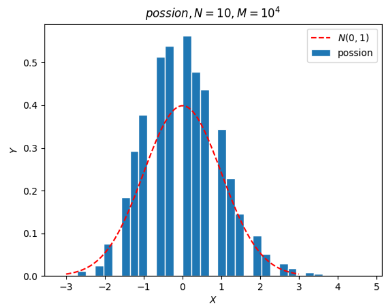

图 10: $M = 10^4$

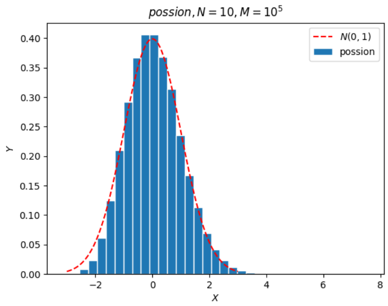

图 11: $M = 10^5$

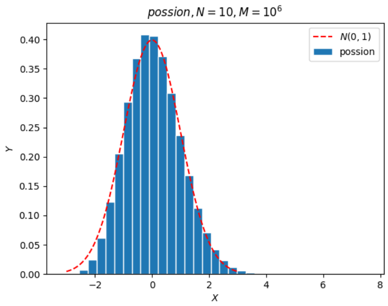

图 12: $M = 10^6$

5.2 指数分布

$N = 2$ 时

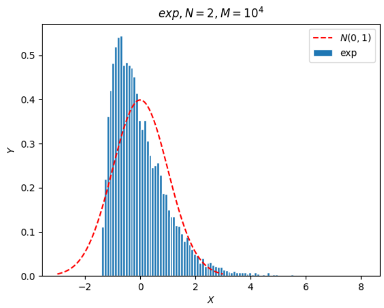

图 13: $M = 10^4$

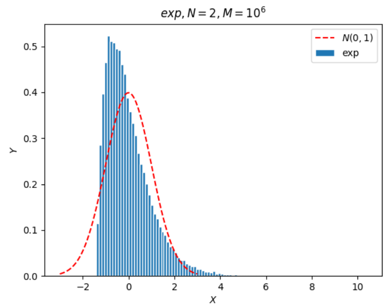

图 14: $M = 10^6$

$N = 5$ 时

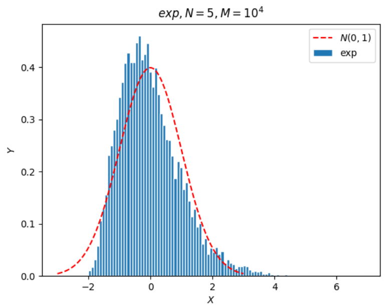

图 15: $M = 10^4$

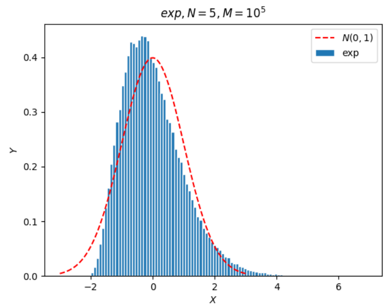

图 16: $M = 10^6$

$N = 10$ 时

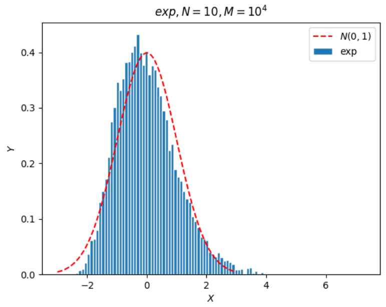

图 17: $M = 10^4$

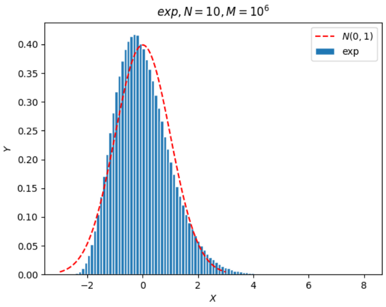

图 18: $M = 10^6$

5.3 两点分布

$N = 2$ 时

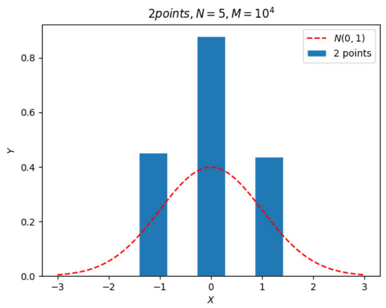

图 19: $M = 10^4$

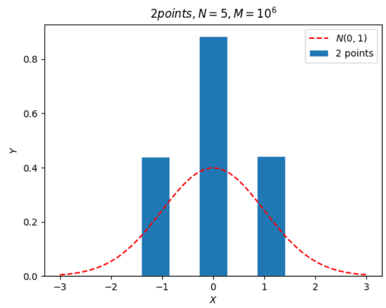

图 20: $M = 10^6$

$N = 5$ 时

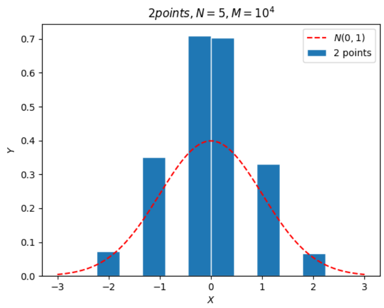

图 21: $M = 10^4$

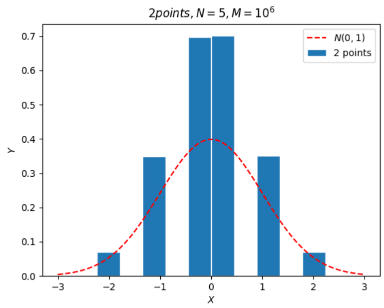

图 22: $M = 10^6$

$N = 10$ 时

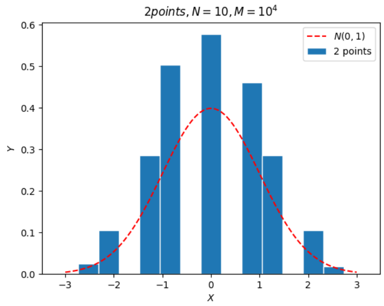

图 23: $M = 10^4$

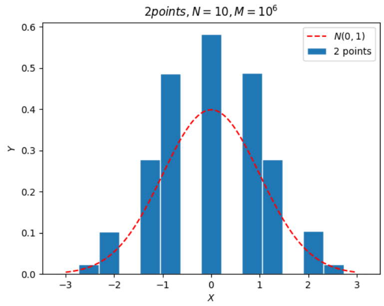

图 24: $M = 10^6$

5.4 均匀分布

$N = 2$ 时

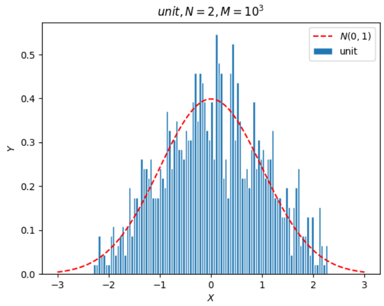

图 25: $M = 10^3$

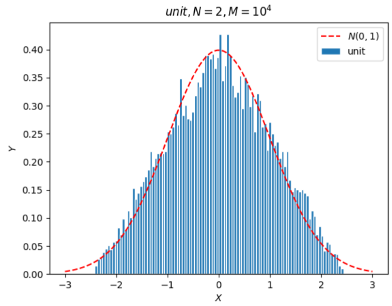

图 26: $M = 10^4$

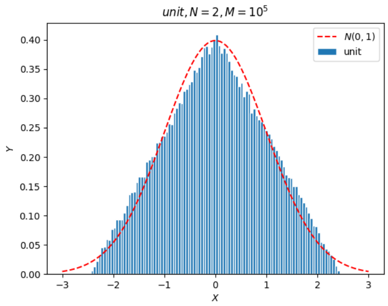

图 27: $M = 10^5$

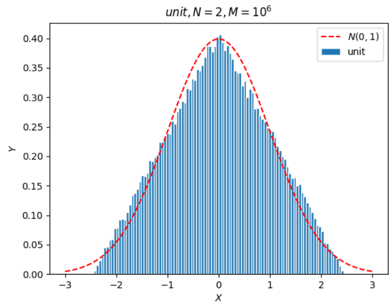

图 28: $M = 10^6$

$N = 5$ 时

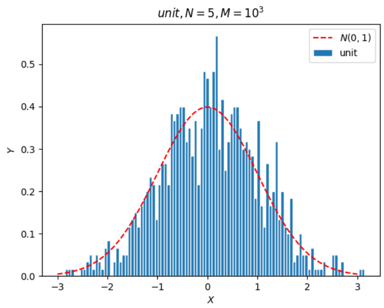

图 29: $M = 10^3$

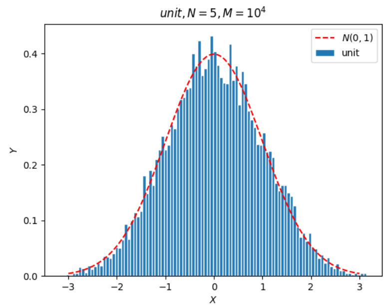

图 30: $M = 10^4$

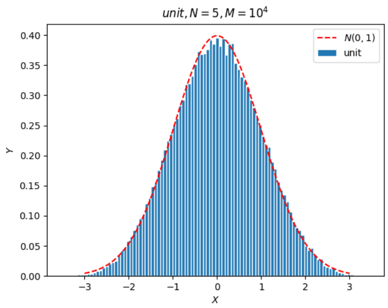

图 31: $M = 10^5$

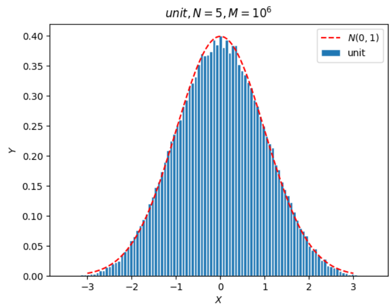

图 32: $M = 10^6$

$N = 10$ 时

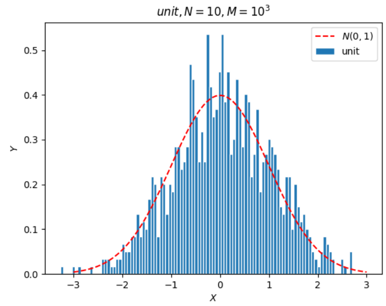

图 33: $M = 10^3$

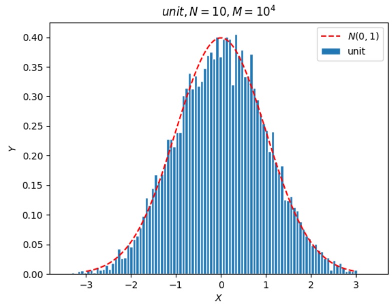

图 34: $M = 10^4$

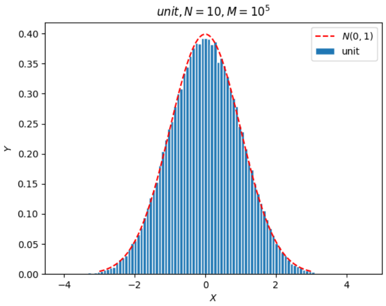

图 35: $M = 10^5$

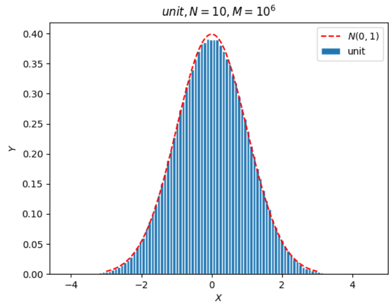

图 36: $M = 10^6$

发现上述的各种分布在 $N = 10, M = 10^6$ 时最接近于标准正态分布。
对于一个相同的 N，当 M 越大的时候，越接近于标准正态分布。
对于一个相同的 M，当 N 越大的时候，越接近于标准正态分布。
所以可以说明中心极限定理对以上的各种分布均成立。

6 总结

1. 对于不同的随机分布，不管是连续分布还是离散分布，均符合中心极限定理
2. 中心极限定理在 N 较大的时候其作用更加明显
3. 中心极限定理在 M 较大的时候其作用更加明显