

# 基于 MATLAB 的 De Moivre – Laplace 中心极限定理的随机模拟

任 丽, 李顺初

(西华大学理学院应用数学研究所, 成都 610039)

**摘 要:**针对 De Moivre – Laplace 中心极限定理, 首先利用特征函数和独立同分布下的中心极限定理得到理论上证明, 然后利用 MATLAB 软件随机模拟分析二项分布与正态分布之间的关系, 并对图像进行误差分析, 最后证明数值结果与实验结果的一致性。通过这两种不同角度的论证, 给出 De Moivre – Laplace 中心极限定理直观的解释和说明, 这种将数学实验与数学原理结合的方法, 不仅使数学原理化抽象为具体, 便于理解和实践, 而且扩展了数学原理和计算机软件的应用。

**关键词:**De Moivre – Laplace 中心极限定理; 正态分布; 二项分布; 随机模拟; 误差分析

**中图分类号:**0021; 0024

**文献标志码:**A

## 引 言

中心极限定理描述了离散型随机变量与连续型随机变量的内在联系, 即离散型随机变量的极限分布是正态分布, 因此中心极限定理的结论使正态分布在概率论和数理统计中具有重要作用<sup>[1]</sup>, 它的研究促进了现代概率论的发展, 并且在统计分析和近似计算具有广泛的应用。近年来, 国内众多学者从两方面对中心极限定理进行了研究: 一方面, 如文献[2 – 9]研究了中心极限定理的实际应用; 另一方面, 如文献[10 – 14]研究了中心极限定理含义和本质, 其研究内容是单一的, 只用了数值方法证明中心极限定理, 并没有详细阐明常见分布(几种离散分布和连续分布)与正态分布的关系, 也没有用数学软件模拟这种关系。

本文以 De Moivre – Laplace 中心极限定理为例, 在蒙特卡罗思想<sup>[15]</sup>下, 研究中心极限定理中二项分布与正态分布的关系, 并通过 MATLAB 软件<sup>[16-20]</sup>对其进行模拟, 直观展现 De Moivre – Laplace 中心极限定理的形成

过程, 并验证数值结果的准确性, 同时对实验结果进行误差分析, 进一步得到影响二项分布近似正态分布的因子, 从而对 De Moivre – Laplace 中心极限定理作出更合理地诠释, 将为中心极限定理在其他分布中的应用提供了一种新的分析手段。这种将计算机和数学原理结合的方法, 不仅将数学原理具体化, 而且还实现了数学实验中计算机软件的应用与扩展。

## 1 预备知识

**引理 1**<sup>[21]</sup> 设  $\{x_n\}$  是独立同分布的随机变量序列, 且  $E(X_i) = \mu, \text{Var}(X_i) = \sigma^2 > 0$ , 若记

$$Y_n^* = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

则对任意实数  $y$ , 有

$$\lim_{n \rightarrow +\infty} P(Y_n^* \leq y) = \Phi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-\frac{t^2}{2}} dt \quad (1)$$

**证明** 设的特征函数  $\varphi(t)$ , 由特征函数的性质可得

收稿日期: 2016-01-03

基金项目: 四川省教育厅自然科学重点项目(12ZJA164)

作者简介: 任 丽(1991-), 女, 四川南充人, 硕士生, 主要从事微分方程及其应用方面的研究, (E-mail)15208416930@163.com;

李顺初(1963-), 男, 湖北黄冈人, 教授, 主要从事微分方程及其应用、油气藏工程、渗流力学方面的研究, (E-mail)lishunchu@163.com

$Y_n^*$  的特征函数为:

$$\varphi_{Y_n^*}(t) = \left[ \varphi\left(\frac{t}{\sigma\sqrt{n}}\right) \right]^n$$

因为

$$E(X_i) = 0, \text{Var}(X_i) = \sigma^2$$

所以有

$$E(X_n - \mu) = 0 = \frac{\varphi'(0)}{i}$$

$$\text{Var}(X_n - \mu) = \sigma^2 = -\varphi''(0) + \varphi''(0)^2$$

由特征函数的性质可得:

$$\varphi'(0) = 0, \varphi''(0) = -\sigma^2$$

又因为特征函数有展开式

$$\varphi(t) = \varphi(0) + \varphi'(0) \cdot t + \frac{\varphi''(0)}{2} \cdot t^2 + o(t^2) =$$

$$1 - \frac{1}{2} \cdot \sigma^2 t^2 + o(t^2)$$

知

$$\lim_{n \rightarrow +\infty} \varphi_{Y_n^*}(t) = \lim_{n \rightarrow +\infty} \left[ 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n^2}\right) \right]^n = e^{-\frac{t^2}{2}}$$

且标准正态分布的特征函数  $\varphi(t) = e^{-\frac{t^2}{2}}$ , 即  $\{Y_n^*\}$  的特征函数数列收敛于标准正态分布的特征函数。

由特征函数的连续性定理可得:  $\{Y_n^*\}$  的分布函数列弱收敛于标准正态分布。

故(1)式得证,引理 1 证明完毕。

只假设  $\{X_n\}$  独立同分布、方差存在,不管原来的分布是什么,只要当  $n$  充分大时,  $Y_n$  才近似服从标准正态分布  $N(0,1)$ , 当  $n$  较小时,此种近似不能保证。所以这个定理有着广泛的应用<sup>[21]</sup>。

## 2 De Moivre - Laplace 中心极限定理证明及模拟算法

### 2.1 De Moivre - Laplace 中心极限定理

设  $n$  重贝努里试验中,事件  $A$  在每次实验中出现的概率为  $p(0 < p < 1)$ , 记  $X_i$  为  $i$  次试验中事件  $A$  出现的次数,即

$$X_i \sim b(1,p) E(X_i) = 0, \text{Var}(X_i) = \sigma^2, i = 1, 2, \dots, n$$

且记

$$Y_n^* = \frac{\sum_{i=1}^n X_i - np}{\sqrt{npq}}$$

则对任意实数  $y$ , 有

$$\lim_{n \rightarrow +\infty} P(Y_n^* \leq y) = \varphi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-\frac{t^2}{2}} dt \quad (2)$$

**证明** 存在  $n$  个相互独立,服从同一  $0-1$  分布的诸随机变量  $X_1, X_2, X_3, \dots, X_n$ , 其中  $X_i (i = 1, 2, \dots, n)$  的分布概率为  $P(X_i = i) = p^i(1-p)^{1-i}, i = 0, 1$ , 因为

$$E(X_i) = p, \text{Var}(X_i) = p(1-p)$$

$$i = 1, 2, \dots, n$$

由林德伯格 - 莱维中心极限定理可得:

$$\lim_{n \rightarrow +\infty} P(Y_n^* \leq y) = \lim_{n \rightarrow +\infty} P\left(\frac{\sum_{i=1}^n X_i - np}{\sqrt{npq}} \leq y\right) =$$

$$\varphi(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-\frac{t^2}{2}} dt$$

### 2.2 De Moivre - Laplace 中心极限定理的随机模拟算法设计

(1) 从计算机中产生 100 个  $b(100, 0.9)$  二项分布的随机数, 记为  $x_1, x_2, x_3, \dots, x_{100}$ 。

(2) 计算

$$y = \frac{x_1 + x_2 + \dots + x_{100} - k \cdot n \cdot p}{\sqrt{k \cdot n \cdot p \cdot (1-p)}}$$

由 De Moivre - Laplace 中心极限定理知, 可将  $y$  近似看成来自标准正态分布  $N(0,1)$  的一个随机数。

(3) 计算  $z = \mu + \sigma y$ , 则可将  $z$  看成来自正态分布  $N(\mu, \sigma^2)$  的一个随机数。

(4) 重复(1) ~ (3)  $k$  次, 就可以得  $N(\mu, \sigma^2)$  分布的  $k$  个随机数。

### 2.3 De Moivre - Laplace 中心极限定理的随机模拟流程图以及程序

De Moivre - Laplace 中心极限定理随机模拟流程图如图 1 所示。

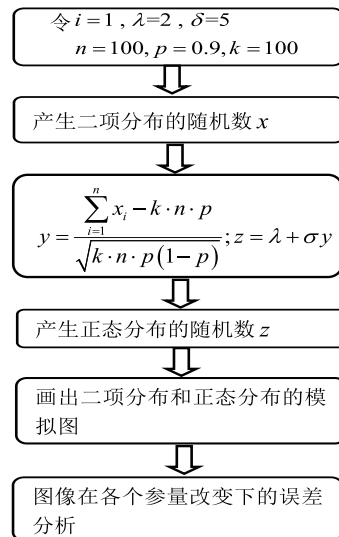


图 1 De Moivre - Laplace 中心极限定理随机模拟流程图

根据流程图1,编写程序如下:

```
clear all;
m = 2000000; % m;运行次数
mu = 2;
sigma = 5;
z = zeros(1, n);
n = 100;
p = 0.9;
k = 100; % 随机数的个数
for i = 1:n
    x = binornd(n, p, 1, k);
    y = (sum(x(1, :)) - k * n * p)
    /sqrt(k * n * p * (1 - p));
    z(i) = mu + sigma * y;
end
zmin = min(z);
zmax = max(z);
t = linspace(zmin, zmax, 100);
zz = hist(z, t);
zz = zz/length(z)/((zmax - zmin)/100);
plot(t, zz, 'r')
hold on;
xxx = -30:0.1:30;
yyy = normpdf(xxx, mu, sigma);
plot(xxx, yyy);
h = legend('二项分布', '正态分布', 1);
for i = 1:k
    e(i) = abs(yyy(i) - zz(i)); % 绝对值误差
end
```

E1 = max(e) % 最大模误差

E2 = norm(e, 2) % 平方误差

在参量  $n = 100, p = 0.9$  下  $m = 200\ 000, m = 2\ 000\ 000, m = 2\ 500\ 000$  的运行结果如图2所示。

#### 2.4 De Moivre - Laplace 中心极限定理随机模拟图象误差分析

要实现 De Moivre - Laplace 中心极限定理的模拟,关键在于尽可能多地进行抽样试验,为了解决这一问题,本文在 MATLAB R2013a 软件环境下产生服从二项分布的随机数代替现实中的抽样试验,由 De Moivre - Laplace 中心极限定理可知:对其随机模拟的运行结果产生影响的变量只有三个,即运行次数  $m$ , 二项分布中  $n$  和  $p$ 。对此,从变量  $m, n, p$  的取值角度出发,讨论对 De Moivre - Laplace 中心极限定理的随机模拟图象误差分

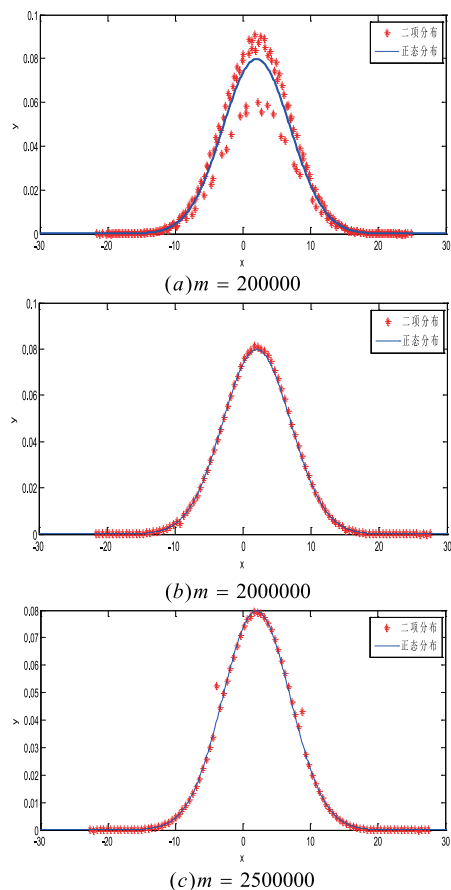


图2 不同运行次数下的运行结果

析,当  $n = 100, p = 0.9$  时改变模拟次数  $m$ , 随机模拟 De Moivre - Laplace 中心极限定理数值分析比较见表1。

表1 不同  $m$  下的误差

模拟结果参数值 (模拟次数 $m$ )	最大模误差	平方误差
$m = 2 \times 10^2$	0.1364	0.4803
$m = 2 \times 10^3$	0.1041	0.3990
$m = 2 \times 10^4$	0.1014	0.3815
$m = 2 \times 10^5$	0.0912	0.3638
$m = 2 \times 10^6$	0.0811	0.3415
$m = 3 \times 10^6$	0.0794	0.3364
$m = 4 \times 10^6$	0.0975	0.3358
$m = 5 \times 10^6$	0.0984	0.3284
$m = 6 \times 10^6$	0.0983	0.3284
$m = 7 \times 10^6$	0.0944	0.3226

由图2可知,随着运行次数的增大,De Moivre - Laplace 中心极限定理随机模拟图像效果愈加显著,二项分布和标准正态分布重合愈紧密;由表1的二项分布与标准正态分布图形之间误差分析也说明了图2结果。

随机模拟运行次数  $m = 2 \times 10^6$ , 在  $n$  取值一定(或  $p$  取值一定),改变  $p$  的取值(或改变  $n$  的取值)随机模拟 De Moivre - Laplace 中心极限定理的数值分析见表2。

表 2 随  $n$  和  $p$  变化的误差分析

模拟结果参数值	最大模误差	平方误差
$n = 100, p = 0.1$	0.0916	0.3683
$n = 100, p = 0.2$	0.0912	0.3643
$n = 100, p = 0.3$	0.0945	0.3556
$n = 100, p = 0.4$	0.0897	0.3542
$n = 100, p = 0.5$	0.0919	0.3652
$n = 100, p = 0.6$	0.0865	0.3519
$n = 100, p = 0.7$	0.0935	0.3524
$n = 100, p = 0.8$	0.0868	0.3529
$n = 100, p = 0.9$	0.0854	0.3517
$n = 1000, p = 0.1$	0.0854	0.3395
$n = 1000, p = 0.2$	0.0828	0.3388
$n = 1000, p = 0.3$	0.0842	0.3434
$n = 1000, p = 0.4$	0.0833	0.3398
$n = 1000, p = 0.5$	0.0807	0.3375
$n = 1000, p = 0.6$	0.0835	0.3422
$n = 1000, p = 0.7$	0.0811	0.3458
$n = 1000, p = 0.8$	0.0841	0.3431
$n = 1000, p = 0.9$	0.0876	0.3453
$n = 1000, p = 0.01$	0.1036	0.3460
$n = 1000, p = 0.02$	0.0993	0.3614
$n = 1000, p = 0.03$	0.0802	0.3505
$n = 1000, p = 0.04$	0.0888	0.3611
$n = 1000, p = 0.05$	0.0869	0.3505
$n = 1000, p = 0.06$	0.0845	0.3575
$n = 1000, p = 0.07$	0.0857	0.3621
$n = 1000, p = 0.08$	0.0908	0.3482
$n = 1000, p = 0.09$	0.0828	0.3467

由表 2 给出的计算误差可知,在随机模拟次数  $m$  和二项分布某个事件发生概率  $p$  一定的情况下,增加进行二项分布样本点的总个数  $n$ ,可减少 De Moivre - Laplace 中心极限定理随机模拟图像的误差;同理,在随机模拟次数  $m$  和进行二项分布样本点的总个数  $n$  一定的情况下,增加二项分布某个事件发生概率  $p$ ,可减少 De Moivre - Laplace 中心极限定理随机模拟图像的误差。但是,也出现某些点的误差随着二项分布样本点的总个数  $n$  和二项分布某个事件发生概率  $p$  的增大反而略有变大的情况,这主要是由于蒙特卡罗方法在本质上属于一种非确定性的方法,并且 De Moivre - Laplace 中心极限定理进行的是随机模拟,从而会显示出一定的随机性。

### 3 结论

(1)在分析 De Moivre - Laplace 中心极限定理的随机模拟图像后发现:随机模拟的运行次数  $m$ ,二项分布中  $n$  和  $p$  这三个因素对于二项分布的正态近似都有一定程度的影响。分别增加运行次数  $m$ ,二项分布中  $n$  和  $p$ ,都可以减少二项分布近似正态分布的误差,但由于蒙特卡罗方法本质上属于一种非确定性方法,因而在分析随机模拟图像某些点的误差时,分别增加运行次数  $m$ 、二

项分布中  $n$  和  $p$  误差反而增大。

(2)基于蒙特卡罗方法的基本思想,通过产生的正态随机数,在 MATLAB 环境下随机模拟中心极限定理,把传统教学中无法实现的大量实验变成现实,从而把原本抽象难懂的知识变得直观具体,易于理解。在计算机的帮助下学习数学知识、分析理论、总结规律,也实现数学实验中计算机软件的应用与扩展,它体现了将数学视为一门“试验科学”的思想。

### 参考文献:

- [1] 唐莉,李雁如.大数定律与中心极限定理的实际应用[J].广东技术师范学院学报,2005(6):75-76.
- [2] 孙大利,王久珂,刘晓阳,等.独立同分布中心极限定理在雨量站网规划中的应用[J].北京大学学报:自然科学版,2015,51(1):35-42.
- [3] 丁健,李红菊.中心极限定理在统计推断中的应用[J].长春师范大学学报,2015,34(2):12-14.
- [4] 陈学慧,赵鲁涛,张志刚.案例式中心极限定理教学研究[J].大学数学,2015,31(2):115-118.
- [5] 孙蓓.中心极限定理及其在若干实际问题中的应用[J].科教导刊,2012(6):65-67.
- [6] 荆江雁.概率方法在计算与证明中的应用[J].常州工学院学报,2011,24(2):30-31.
- [7] 徐姿奕,汪四水.独立同分布中心极限定理的应用[J].山西大同大学学报:自然科学版,2007,23(1):13-14.
- [8] 杜伟娟,于文娟.中心极限定理及其初步应用[J].内蒙古电大学刊,2007(7):107-108.
- [9] 岳金健.利用大数定律和中心极限定理求解极限[J].龙岩学院学报,2007,25(3):95-97.
- [10] 石峰利.几个常见统计模拟问题及其在 excel 中的实现和动态展示[D].云南:云南师范大学,2013.
- [11] 蒋观敏.Mathcad 在概率统计模拟实验教学中的应用[D].重庆:西南大学,2011.
- [12] 农吉夫,黄文宁.随机模拟实验在概率极限理论教学中的应用[J].广西民族大学学报:自然科学版,2008,14(1):101-104.
- [13] 罗弟亚.大数定理和中心极限定理可视化[J].四川工业学院学报,2005(6):70-75.
- [14] 林小革,吴文杰.用 MATLAB 模拟大数定律和中心极限定理[J].运城高专学报,1990(3):12-18.
- [15] 文军,严忠权.基于 matlab 的离散事件的随机模拟

- [J].黔南民族师范学院学报,2011(6):17-23.
- [16] 刘泽显,刘志伟,王勤龙.基于 Matlab 的概率统计实验教学的探索与实践[J].贺州学院学报,2014,30(4):122-126.
- [17] 庄光明,夏建伟,彭作祥,等.基于 Matlab 的 Poisson 分布随机数的 Monte carlo 模拟[J].数学的实践与认识,2012,42(5):87-92.
- [18] 农吉夫,黄文宁.随机模拟实验在概率极限理论教学中的应用[J].广西民族大学学报,2008,14(1):101-104.
- [19] 张杰.MATLAB 在计算物理课程教学中的应用[J].计算机应用与软件,2005,22(6):131-132.
- [20] 马锋,封盛.在工程风险分析中应用 MATLAB 进行蒙特卡洛模拟[J].经济管理,2005,25(2):156-157.
- [21] 茆诗松,程依明,濮晓龙.概率论与数理统计教程[M].2版.北京:高等教育出版社,2011.

## Random Simulation of De Moivre-Laplace Central Limit Theorem

REN Li, LI Shunchu

(Institute of Applied Mathematics, College of Science, Xihua University, Chengdu 610039, China)

**Abstract:** Based on De Moivre-Laplace Central Limit Theorem, first, this paper gives a theoretical proof on the theorem by characteristic function and Lindeberg-Levy Theorem. Second, this paper explains and simulates the relationship between normal distribution and binomial distribution by using MATLAB software, and then analyzes the error for the image that is produced by the MATLAB software. Last, prove that the numerical result and the analysis result are consistent. By these two different angles of argument, this paper gives the intuitive explanation and illustration of Moivre-Laplace De central limit theorem. This method combined mathematical experiment and mathematical principles, not only makes mathematical principles specific, easy to understand and practice, but also extends the application of the mathematical principles and computer software.

**Key words:** De Moivre-Laplace Central Limit Theorem; normal distribution; binomial distribution; random simulation; error analysis