探索极限状态下的概率论之美

——《概率论与数理统计》课后 MATLAB 自主实验

西北工业大学航空学院　　课程老师：王燕萍老师

01011704 班　　冯铮浩　　学号：2017300281　　完成时间：2018 年 12 月 3 日

一、实验背景

在本学期《概率论与数理统计》的课程学习中，许多有关极限状态的概率论结论非常抽象又十分有趣。同时，我也发现，这些奇妙的结论在解释实际生活中各类概率问题，以及实际工程近似与简化计算方面起到了举足轻重的作用，大大拓展了概率论的应用方式。

例如，泊松定理指出，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，从而给出一种简洁的二项分布概率的近似计算方式，广泛应用于实际工程计算。另外，第五章涉及的两类著名的极限定理中，“大数定律”证明了频率具有稳定性，即当试验次数不断增大时，频率稳定在一个数的附近，很好地解释了人们在长期实践中的许多经验原理。“中心极限定理”表明，在相当一般的条件下，当独立随机变量的个数不断增加时，其和的分布趋向于正态分布，揭示了为什么在实际应用中会经常遇到正态分布，也即揭示了生活中产生如此多正态分布变量的源泉。

但是，这些包含“极限”与“趋向于”概念的睿智定律或定理的纯理论证明比较抽象，涉及概念也较多较复杂。由此，通过老师课堂传授与前期自己对相关数学软件知识的学习，我构想利用 MATLAB 软件对前面提到的**泊松定理**、**大数定律**和**中心极限定理**进行随机仿真模拟，并通过图形将模拟结果直观、动态地演示出来，从而对这两个定理所反映的本质内容给出直观的解释和说明，进一步感受极限状态下的概率论规律之美，并提升自己运用知识解决问题的能力。

二、理论准备

1. 泊松定理 设 $\lambda > 0$ 是一个常数，n 是任意正整数，设 $np_n = \lambda$，则对于任一固定的非负整数 k，有

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

2. 弱大数定理（辛钦大数定理） 设 $X_1, X_2, \cdots$ 是相互独立，服从同一分布的随机变量序列，且具有数学期望 $E(X_k)=\mu(k=1,2,\cdots)$.作前 n 个变量的算术平均 $\frac{1}{n}\sum_{k=1}^{n} X_k$，则对于任意 $\varepsilon > 0$，有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| < \varepsilon\right\} = 1. \tag{2}$$

3. 独立同分布的中心极限定理 设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，服从同一分布，且具有数学期望和方差：$E(X_k)=\mu, D(X_k)=\sigma^2>0(k=1,2,\cdots)$，则随机变量之和 $\sum_{k=1}^{n} X_k$ 的标准化变量

$$Y_n = \frac{\sum_{k=1}^{n} X_k - E(\sum_{k=1}^{n} X_k)}{\sqrt{D\left(\sum_{k=1}^{n} X_k\right)}} = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}$$

的分布函数 $F_n(x)$ 对于任意 x 满足

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\left\{\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \le x\right\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x). \tag{3}$$

即当 n 充分大时，有

$$\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \overset{\text{近似地}}{\sim} N(0,1). \tag{4}$$

注：其他各种大数定律与中心极限定理在这里没有详细列出。

4. 对于经验分布函数 $F_n(x)$，**格里汶科(Glivenko)**在 1933 年证明了以下的结果：对于任一实数 x，当 $n\to\infty$ 时 $F_n(x)$ 以概率 1 一致收敛与分布函数 $F(x)$，即

$$P\left\{\lim_{n\to\infty}\sup_{-\infty<x<\infty}\left|F_n(x)-F(x)\right|=0\right\}=1. \tag{5}$$

因此，对于任一实数 x，当 n 充分大时，经验分布函数的任一个观察值 $F_n(x)$ 与总体分布函数 $F(x)$ 只有微小的差别，从而在实际上可当作 $F(x)$ 来使用。

三、使用软件

MATLAB R2017a (64 位,Window 10 下运行)

四、程序实现与结果可视化

1. 泊松定理——MATLAB 仿真模拟验证

A. 实现方式与思路

泊松定理涉及泊松分布、二项分布与极限思想。我的具体实现思路如下：

① 首先，由定理的条件 $np_n=\lambda$（常数），给出一个 λ 的确定值作为泊松分布的参数；

② 接着，设置适当的 n 值，满足从较小值到较大值离散变化；

③ 然后，利用 binopdf 与 poisspdf 两个标准作图指令，保持泊松分布的图线不变，作出不同的 n 值对应的二项分布图线；

④ 最后，利用 MATLAB 指令对图像进行修饰，并观察图像变化规律。

B. 实际参数设置与绘制图线结果

在这里，我设置泊松分布的参数 $\lambda=100$（常数），$n=300,1000,1500,10000$，满足取值中包括较小值、过渡值与较大值，从而能够反映图像变化的趋势。作图结果如下图 1,2,3,4 所示。

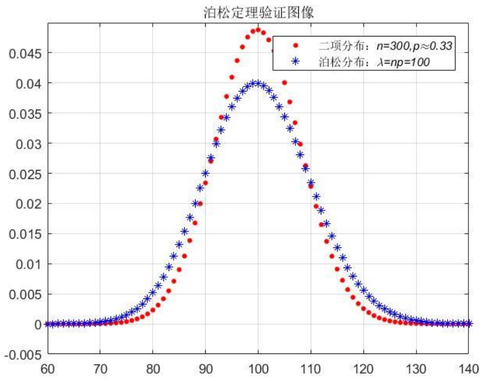

图 1 分布图线 $(n=300, \lambda=100)$

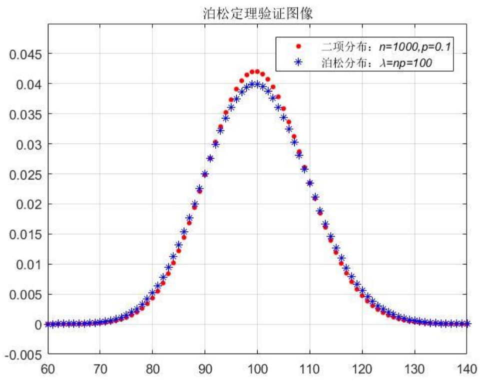

图 2 分布图线 $(n=1000, \lambda=100)$

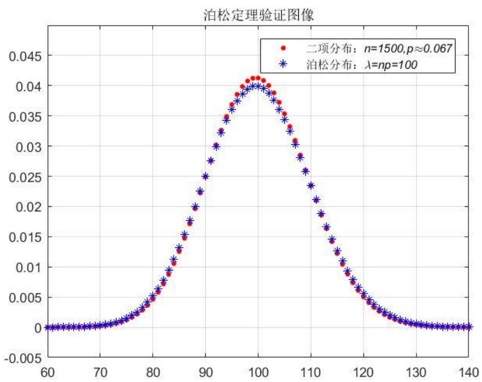

图 3 分布图线 $(n=1500, \lambda=100)$

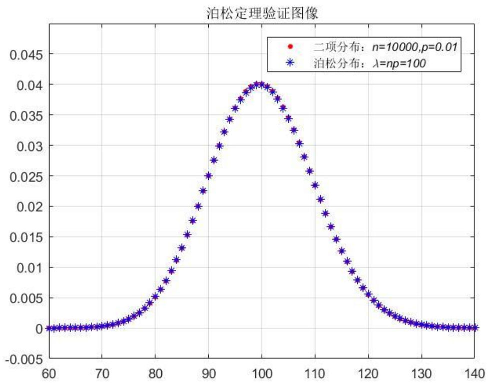

图 4 分布图线 $(n=10000, \lambda=100)$

C. 结果分析

从 MATLAB 绘制的四张分布曲线点图可以清晰看出，当 n 增大时，二项分布的数据点曲线向泊松分布数据点曲线逼近。当 n=300（较小值）时，两者有明显差距；而当 n=10000（较大值）时，两组数据点几乎重合，进一步印证了当 $n \to \infty$ 时，我们可以用泊松分布逼近二项分布，大大简化计算量。

2. 辛钦大数定理——MATLAB 仿真模拟验证

A. 实现方式与思路

为了使用计算机仿真辛钦大数定理的极限实现过程，由格里汶科(Glivenko)给出的结论，我将大量独立随机变量的算术平均 $\frac{1}{n}\sum_{k=1}^{n} X_k$ 看作总体[1]，并对总体进行抽样，通过研究抽样样本的分布来实现对总体本身分布的推断。这里，我利用 MATLAB 生成的服从某一分布的一组随机数来模拟实际的一次随机实验，算法

的关键在于实现尽可能多的抽样实验。

具体算法流程图如下图 5 所示，

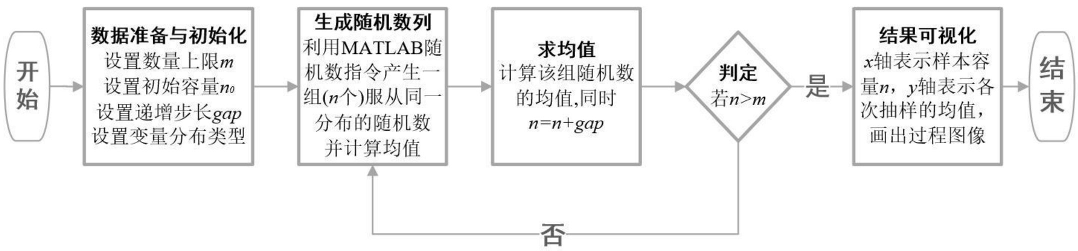

图 5 辛钦大数定理仿真验证算法 MATLAB 实现流程图

B. 实际参数设置与绘制图线结果

在这里，我设置随机变量均服从泊松分布，且参数 $\lambda = 100$，设置随机变量数量上限 $m = 10000$，初始样本容量 $n_0 = 5$，递增步长 $gap = 5$，作出的图线如下图 6 所示。

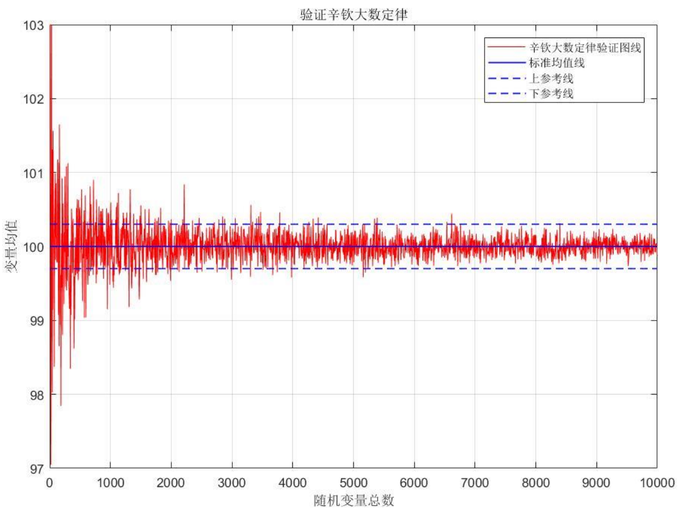

图 6 辛钦大数定理 MATLAB 仿真验证图线

C. 结果分析

从图形变化来看，随着随机变量总数的递增，$\frac{1}{n}\sum_{k=1}^{n} X_k$ 的取值点逐渐密集于总体分布期望 100 的附近，当总数>4000 时，随机模拟曲线几乎都位于带状区域 100 ± 0.3 内，这很好地反映了辛钦大数定理反映的统计规律性。

D. 拓展思考

在第一步中，若将泊松分布换为 0-1 分布，则算法模拟的是辛钦大数定理的一个重要推论，即伯努利大数定理。

3. 独立同分布的中心极限定理——MATLAB 仿真模拟验证

A. 实现方式与思路

与辛钦大数定理类似，独立同分布的中心极限定理的仿真实现需要尽可能多地实现样本抽样，将$\sum_{k=1}^{n} X_k$看成总体，设$X_i = \dfrac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}$

具体算法流程图如下图 7 所示。

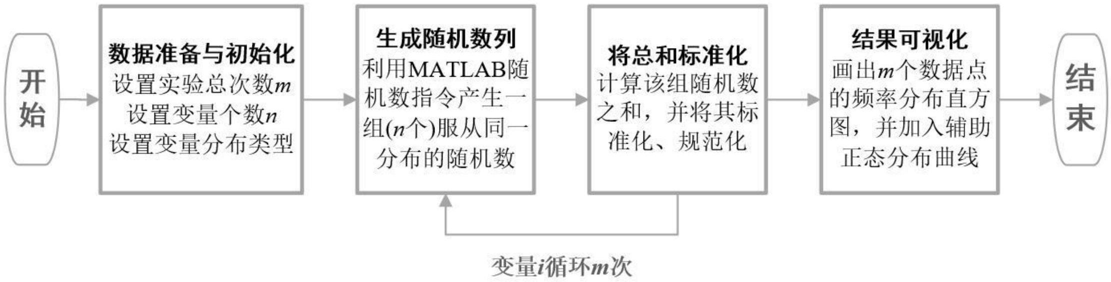

图 7 中心极限定理仿真验证算法 MATLAB 实现流程图

B. 实际参数设置与绘制图线结果

在这里，我设置随机变量均服从参数$\theta = 4$的指数分布，即$X \sim \mathrm{Exp}(4)$，数学期望为 4，方差为 16。设置实验总次数$m = 10000$，变量个数满足$n = 10,100,1000$，作出的三张频率分布直方图如下图 8,9,10 所示。

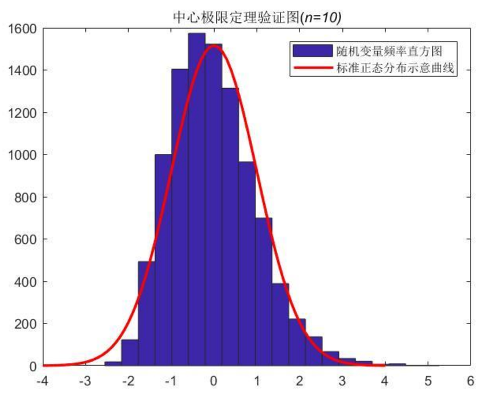

图 8 频率分布直方图$(n = 10, \theta = 4)$

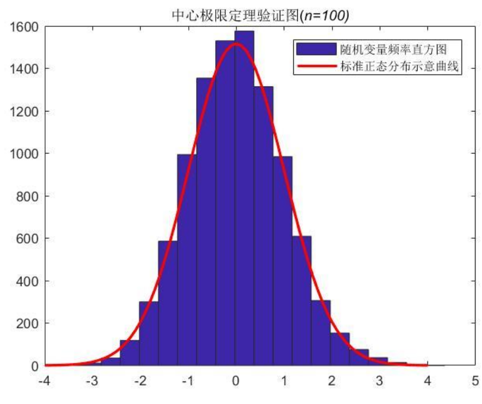

图 9 频率分布直方图$(n = 100, \theta = 4)$

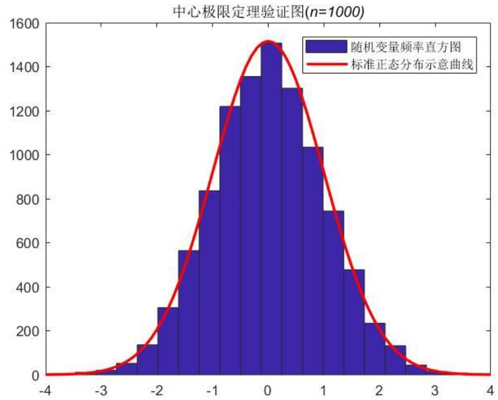

图 10 频率分布直方图$(n=1000, \theta=4)$

C. 结果分析

为了使观察结果更加便捷、清晰，我在每张频率分布直方图中加入了一条辅助曲线，该曲线满足正态分布，方差$\sigma^2=1$。从三张n取不同值的分布直方图中，可以明显看出小长方形中间高，两边低，且总体的高度分布越来越趋向于正态分布。这种变化的过程充分检验了独立同分布的中心极限定理的正确性。

D. 拓展思考

若在第一步中，若将指数分布换为 0-1 分布，则算法模拟的是独立同分布的中心极限定理的一个特殊情况，即棣莫弗-拉普拉斯(De Moivre-Laplace)定理。这个定理表明，正态分布是二项分布的极限分布。

五、总结与思考

在本文中，我结合《概率论与数理统计》课本上有关概率的极限定理理论，运用 MATLAB 软件编程对定理中涉及到的$n\to\infty$情况进行仿真模拟，并利用 MATLAB 强大的图形可视化能力，将定理模拟结果用曲线、频率分布直方图等多种形式进行展示，一方面充分验证了定理的正确性，另一方面我对于这些前人总结的充满智慧的规律也有了更加直观与深刻的理解。但是，实验内容细节还不够充分，方法还不够多元。在今后的学习生活中，我也会更加注重将课堂学习与亲身实验相结合，培养能力与兴趣，为未来的科研竞赛之路打下更加坚实的基础！

最后，感谢王老师在课堂上的教诲，启发我们运用数学软件工具来进行实践，感受更多有关概率论与数理统计的精彩。希望老师批评指正！

六、参考文献

[1] 林小苹,吴文杰.用 MATLAB 模拟大数定律和中心极限定理[J].汕头大学学报(自然科学版),2005(02):12-18+80.

七、附录（附本文涉及全部 MATLAB 代码）

%% 程序一文件名：Possion_Rules.m

```
%% 泊松定理MATLAB程序展示——图形可视化
%% 数据准备
clear all
clc
lambda=100; % 泊松分布参数(常数100)
%% 绘制二项分布、泊松分布分布律图(对应不同n值)
%% n=300（较小值）
figure;
n=300;     % 二项分布中的n
p=lambda/n;     % 二项分布中的p
hold on;
x=0:n;
y1=binopdf(x,n,p);
y2=poisspdf(x,lambda);
plot(x,y1,'r.','MarkerSize',12);
plot(x,y2,'b*','MarkerSize',6);
legend('二项分布：\itn=300,\itp\approx0.33','泊松分布：\it\lambda=\itnp=100'); % 显示图例
axis([lambda-40,lambda+40,-0.005,0.05]); % 调整坐标轴
title('泊松定理验证图像');
grid on;
box on;

%% n=1000
figure;
n=1000;     % 二项分布中的n
p=lambda/n;     % 二项分布中的p
hold on;
x=0:n;
y1=binopdf(x,n,p);
y2=poisspdf(x,lambda);
plot(x,y1,'r.','MarkerSize',12);
plot(x,y2,'b*','MarkerSize',6);
legend('二项分布：\itn=1000,\itp=0.1','泊松分布：\it\lambda=\itnp=100');    % 显示图例
axis([lambda-40,lambda+40,-0.005,0.05]);    % 调整坐标轴
title('泊松定理验证图像');
```

```
grid on;
box on;

%% n=1500
figure;
n=1500;    % 二项分布中的n
p=lambda/n;    % 二项分布中的p
hold on;
x=0:n;
y1=binopdf(x,n,p);
y2=poisspdf(x,lambda);
plot(x,y1,'r.','MarkerSize',12);
plot(x,y2,'b*','MarkerSize',6);
legend('二项分布：\itn=1500,\itp\approx0.067','泊松分布：\it\lambda=\itnp=100');  % 显示图例
axis([lambda-40,lambda+40,-0.005,0.05]);    % 调整坐标轴
title('泊松定理验证图像');
grid on;
box on;

%% n=10000（较大值）
figure;
n=10000;    % 二项分布中的n
p=lambda/n;    % 二项分布中的p
hold on;
x=0:n;
y1=binopdf(x,n,p);
y2=poisspdf(x,lambda);
plot(x,y1,'r.','MarkerSize',12);
plot(x,y2,'b*','MarkerSize',6);
legend('二项分布：\itn=10000,\itp=0.01','泊松分布：\it\lambda=\itnp=100');  % 显示图例
axis([lambda-40,lambda+40,-0.005,0.05]);    % 调整坐标轴
title('泊松定理验证图像');
grid on;
box on;
```

%% 程序二文件名：Law_of_Large_Numbers.m

```
%% 大数定律的MATLAB验证——图形可视化
%% 数据初始化
clear all
clf
clc
nmax=10000;  % 设置随机变量数量上限
n=5; % 设置初始样本容量
```

```
gap=5; %  设置步长
tot=0;
x=zeros(1,nmax/n);
y=zeros(1,nmax/n);
while n<=nmax
    d=poissrnd(100,1,n); %  设置样本服从参数为100的泊松分布
    E=mean(d);
    tot=tot+1;
    x(tot)=n;
    y(tot)=E;
    n=n+gap;
end

%%  画图表示
xx=0:nmax;
y_mu_line=100*ones(1,length(xx));
y_up_line=100.3*ones(1,length(xx));
y_down_line=99.7*ones(1,length(xx));
hold on;
plot(x,y,'r-','LineWidth',0.5);
plot(xx,y_mu_line,'b-','LineWidth',1);
plot(xx,y_up_line,'b--','LineWidth',1);
plot(xx,y_down_line,'b--','LineWidth',1);
%  调整图像
legend('辛钦大数定律验证图线','标准均值线','上参考线','下参考线');
axis([0,nmax,97,103]);
title('验证辛钦大数定律');
xlabel('随机变量总数'),ylabel('变量均值');
grid on;
box on;
```

%% 程序三文件名：Central_Limit_Theorem.m

```
%%  中心极限定理的MATLAB验证(n=10,100,1000)——图形可视化
%%  数据准备与初始化
clear all
clf
clc
m=10000;    %  设置总体样本容量
%%  随机变量数目为10
n=10; %  设置随机变量个数
e=4; %  设置指数分布参数
tot=0;
mt=zeros(1,m);
for i=1:m
```

```
    d=exprnd(e,1,n); % 设置样本服从参数为e的指数分布
    s=sum(d);
    tt=(s-n*e)/(sqrt(n)*e); % 变量规范化
    tot=tot+1;
    mt(tot)=tt;
end
% 画图表示
hist(mt,20); % 画出频率分布直方图
x1=-4:0.1:4;
y1=normpdf(x1,0,1);
y1=y1*3800;
hold on
plot(x1,y1,'r-','LineWidth',2);
title('中心极限定理验证图(\itn=10)');
legend('随机变量频率直方图','标准正态分布示意曲线');

%% 随机变量数目为100
n=100; % 设置随机变量个数
e=4; % 设置指数分布参数
tot=0;
mt=zeros(1,m);
for i=1:m
    d=exprnd(e,1,n); % 设置样本服从参数为e的指数分布
    s=sum(d);
    tt=(s-n*e)/(sqrt(n)*e); % 变量规范化
    tot=tot+1;
    mt(tot)=tt;
end
% 画图表示
figure;
hist(mt,20); % 画出频率分布直方图
x1=-4:0.1:4;
y1=normpdf(x1,0,1);
y1=y1*3800;
hold on
plot(x1,y1,'r-','LineWidth',2);
title('中心极限定理验证图(\itn=100)');
legend('随机变量频率直方图','标准正态分布示意曲线');

%% 随机变量数目为1000
n=1000; % 设置随机变量个数
e=4; % 设置指数分布参数
tot=0;
mt=zeros(1,m);
```

```
for i=1:m
        d=exprnd(e,1,n); % 设置样本服从参数为e的指数分布
        s=sum(d);
        tt=(s-n*e)/(sqrt(n)*e); % 变量规范化
        tot=tot+1;
        mt(tot)=tt;
end
% 画图表示
figure;
hist(mt,20); % 画出频率分布直方图
x1=-4:0.1:4;
y1=normpdf(x1,0,1);
y1=y1*3800;
hold on
plot(x1,y1,'r-','LineWidth',2);
title('中心极限定理验证图(\itn=1000)');
legend('随机变量频率直方图','标准正态分布示意曲线');
```