2018-2019概率论上机实验题

姓名 何昕圆

班级 能动A72

学号 2173611856

2018-2019概率论上机实验题

何昕圆 能动A72 2173611856

第1题：对二项分布事件概率的精确计算与用泊松分布和中心极限定理的近似计算进行对比。分（a）p变化，n固定，进行比较。（2）p固定,n变化，进行比较。

解：

Matlab代码：

n=10000;

p=0.6;

%由于此题要比较三类方式，第三类为中心极限定理法。故应选取累积概率（分布函数）研究

%对成功次数小于等于6000次总概率进行研究

%在n=10000不变，p改变的情况下

p1=0.1:0.01:0.9;%p值变化范围

P1=binocdf(6000,10000,p1);

subplot(2,1,1)

plot(p1,P1,'k')

hold on;

P2=poisscdf(6000,10000\*p1);

plot(p1,P2,'b')

hold on;

P3=normcdf(6000,10000\*p1,10000\*p1.\*(1-p1));

plot(p1,P3,'r')

xlabel('成功概率p')

ylabel('小于等于60次累计概率P ')

title('p变化n固定比较图')

text(0.56,0.9,'二项')

text(0.5,0.85,'泊松')

text(0.3,0.8,'中心极限定理')

hold off;

%在p=0.6不变，n改变的情况下

n1=100:1:300;%n值变化范围

P1=binocdf(60,n1,0.6);

subplot(2,1,2)

plot(n1,P1,'k')

hold on;

P2=poisscdf(60,n1\*0.6);

plot(n1,P2,'b')

hold on;

P3=normcdf(60,n1\*0.6,0.24\*n1);

plot(n1,P3,'r')

xlabel('实验次数n')

ylabel('小于等于60次累计概率P ')

title('n变化p固定比较图')

text(105,0.1,'二项')

text(120,0.15,'泊松')

text(140,0.3,'中心极限定理')

hold off;

运行结果：

1. 画出的对比图线：



总结：

1. 对于固定n而改变p时，由数据图形对比发现随着概率P的改变，泊松分布的计算值与二项式分布的计算值近似拟合，即当p较小时，泊松分布与二项式分布的近似拟合度增大，但无论P为多大，使用中心极限定理得到的近似计算结果与真实结果相差较大，故中心极限定理不适合二项式分布的拟合。
2. 对于固定P而改变n时，随着数据样本容量n逐渐增大时，二项式分布与泊松分布拟合的程度增大，并且仅当样本容量n非常大时，才能近似服从期望为nμ,方差为nσ^2的正态分布，即当样本容量n特别大时，中心极限才能近似成立。
3. 故当n较大，P较小时使用泊松分布对二项式分布的拟合更为精确。

第2题：

对正态总体参数的区间估计，进行验证及区间长度的变化情况（注：对一个参数，验证一种情形即可）。分（a）样本容量固定，置信度变化；（b）置信度固定，样本容量变化。

解：

Matlab代码：

%利用课本上的一串从正态分布总体取样的数据实验验证区间估计,

E=[12.15 12.12 12.01 12.08 12.09 12.16 12.03 12.01 12.06 12.13 12.07 12.11 12.08 12.01 12.03 12.06];

[muhat,sigmahat,muci,sigmaci]=normfit(E,0.05)%muci是均值的区间估计，sigmaci是方差区间估计

%选取正态分布参数均值，对其进行双侧区间估计

%分两种情况对区间长度研究：

% （a）样本容量固定为16，置信度变化

E=[12.15 12.12 12.01 12.08 12.09 12.16 12.03 12.01 12.06 12.13 12.07 12.11 12.08 12.01 12.03 12.06];

Eb=sum(E)/16;

Ed=0;

for i=1:16

Ed=(E(i)-Eb)^2+Ed;

end

Ed=Ed/15;

%这里n=16，均值为Eb，方差为Ed；

a=[0.25 0.10 0.05 0.025 0.01 0.005]\*2;

a1=1-a;

D=zeros(1,5);

for i=1:6

D(i)=-2\*sqrt(Ed)/sqrt(16)\*tinv(a(i)/2,15);

end

subplot(2,1,1)

plot(a1,D)

xlabel("置信度变化");

ylabel("置信区间长度");

title('置信度变化对区间长度变化影响')

%（b）置信度固定，样本容量变化

%置信度设定为99% 即a=0.01

n=16;

E2=[12.15 12.12 12.01 12.08 12.09 12.16 12.03 12.01 12.06 12.13 12.07 12.11 12.08 12.01 12.03 12.06];

Eb2=sum(E)/16;

Ed2=0;

D2=zeros(1,5);

for n=12:16

for i=1:n

Ed2=(E2(i)-Eb2)^2+Ed2;

end

Ed2=Ed2/(n-1);

D2(n-11)=-2\*sqrt(Ed2)/sqrt(n-1)\*tinv(0.01/2,n-1);

end

subplot(2,1,2)

plot(12:16,D2)

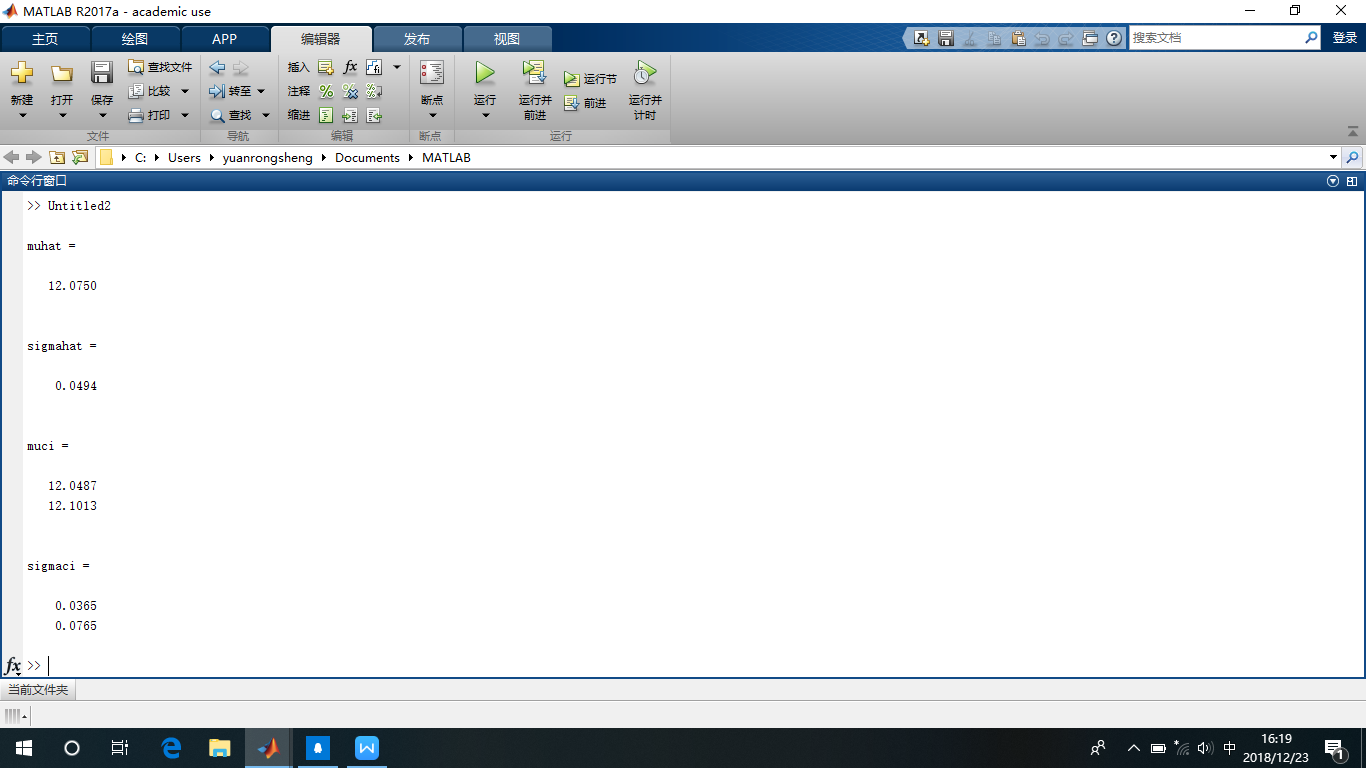
xlabel("样本数变化");

ylabel("置信区间长度");

title('样本数变化对区间长度变化影响')

运行结果：

1. 对区间估计法的验证：



1. 选取正态分布参数均值，对其进行双侧区间估计，探究不同情况区间长度变化：



总结：

1. 对正态总体参数的估计时，固定样本数，置信区间长度会随着置信度的增加而增大，置信区间的置信度越高，则精度越低，反之，若精度越高，置信度越低。
2. 若固定置信度，若样本数增大，则估计量的置信长度减少，即估计量的精度增大，即在实际参量的估计中，如若选取更多的样本数，则估计量的置信区间长度会下降，其精度会增大。

第3题：自己选一个总体，验证样本k阶矩的观察值随样本容量的增大与总体k阶矩接近程度。（对k=1,2进行验证）

解：

Matlab代码：

%随机生成正态分布样本

E=(normrnd(160,25,1,120))

g=size(E,2)

Eb=sum(E)/g;%总体一阶原点矩

Eb2=sum(E.^2)/g;%总体二阶原点矩

Ed=0;%总体一阶中心距

Ed2=var(E)\*(g-1)/g;%总体二阶中心距

D1=zeros(1,g/2);%yangben yi jie yuandianju

D2=zeros(1,g/2);%yangben er jie yuandianju

D3=zeros(1,g/2);%yangben yi jie zhongxinju

D4=zeros(1,g/2);%yangben er jie zhongxinju

for n=g/2:g

F=zeros(1,n);

for i=1:n

F(i)=E(i);

end

Xf=sum(F)/n;

D1(n-g/2+1)=Xf;

D2(n-g/2+1)=sum(F.^2)/n;

D3(n-g/2+1)=0;

D4(n-g/2+1)=var(F)\*(n-1)/n;

end

subplot(2,2,1)

plot(g/2:g,D1)

hold on;

plot(g/2:g,Eb\*ones(1,g/2+1))

%axis([12 16 12.060 12.090])

title("一阶原点矩")

xlabel("样本数");

hold off

subplot(2,2,2)

plot(g/2:g,D2)

hold on;

plot(g/2:g,Eb2\*ones(1,g/2+1))

%axis([12 16 145.6 146.5])

title("二阶原点矩")

xlabel("样本数");

hold off

subplot(2,2,3)

plot(g/2:g,D3)

hold on;

plot(g/2:g,Ed\*ones(1,g/2+1))

title("一阶中心矩")

xlabel("样本数");

hold off

subplot(2,2,4)

plot(g/2:g,D4)

hold on;

plot(g/2:g,Ed2\*ones(1,g/2+1))

title("二阶中心矩")

xlabel("样本数");

hold off

运行结果：（蓝线样本，红线总体）

、

总结：

通过实验数据验证，发现随着样本容量n的增大，样本的k阶中心矩与k阶原点矩分别与总体的k阶中心矩和k阶原点矩的接近程度增大，即若增大样本容量数n，样本k阶矩对拟合总体的k阶矩的拟合程度增大，因此用矩估计法对参数的估计量的准确程度随着样本容量的增大而增加。

第4题：自己设计一种情形，当样本至少为多少时，产品的合格率才能符合给定的合格率。

解:

Matlab代码：

%选取总体服从p=0.95的二项分布,即总体给定的合格率为0.95,样本容量为n

n=1;s=0;i=1;

while i==n

x=binornd(1,0.95,1,1)

if x==1

s=s+1;

end

p=s/n;

if p==0.95

break;

end

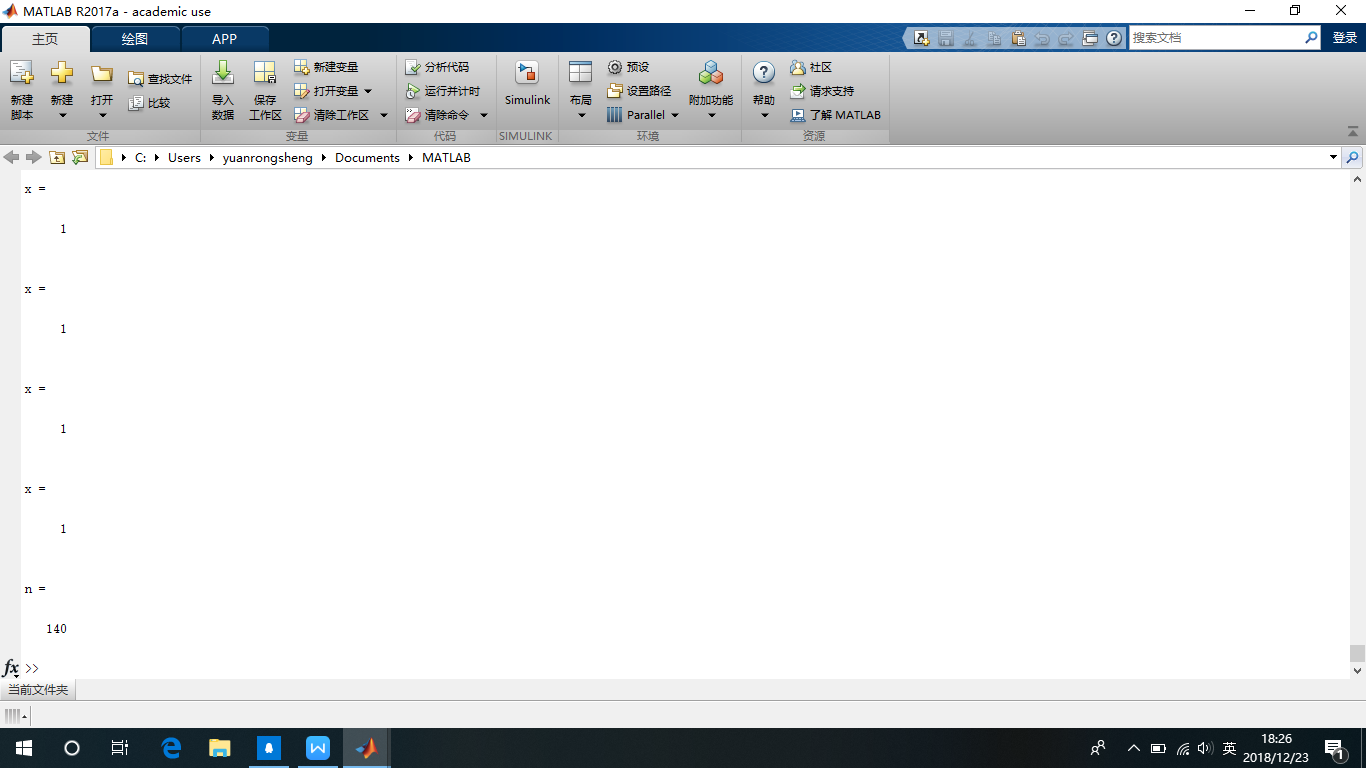
n=n+1;

i=i+1;

end

n=n %此结果含义即为样本所需容量

运行结果：



总结：

本题给定样本的合格率为0.95，通过上机实验发现，当取样次数足够大（至少为140时），才能认定产品的合格率符合给定的合格率。当样本数越接近设定的总体的数量时，越接近设定的合格率。

第5题：画一个分布的统计直方图。

解：

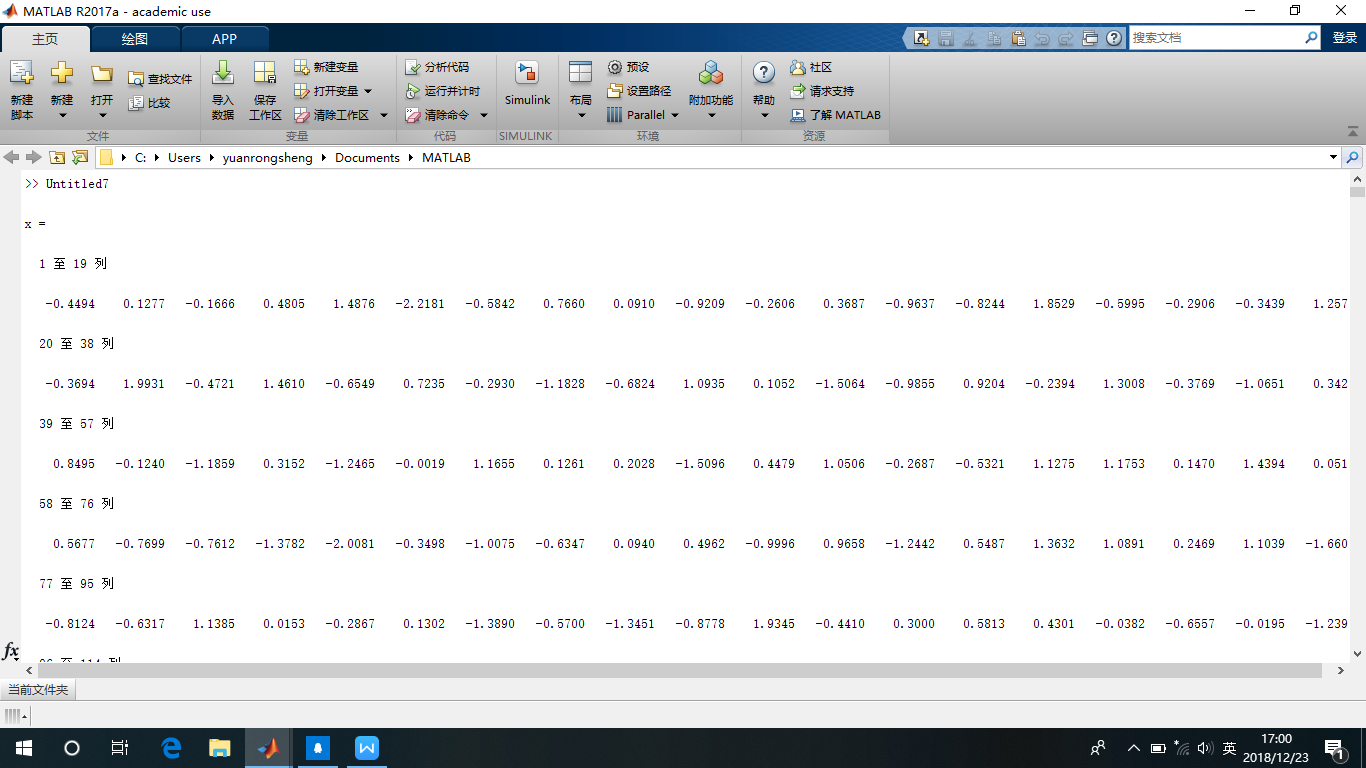
Matlab代码：

x=randn(1,5000)%返回一组5000个的服从标准正态分布的随机项

hist(x,30)%将5000个数据分组绘制成30个直方

运行结果：

（1）产生的5000个数据截图：



注：为了画出的直方图效果更接近标准正态分布曲线，使用数据太多，截取部分数据，无法全部截屏下来展示

1. 画出的直方图：



总结：

本题思想是从一个标准正态总体中随机选出5000个样本，并将样本按数值大小分为30组，每组的面积即为处于该数值范围内的样本个数，故其直方图的高度与总体量的比值可近似代替在该样本数值下的概率密度。由实验得到的直方图可知，样本分布近似于标准正态分布，可用来验证切比雪夫大数定律。

对概率统计学科的感想体会：

通过这一个学期以来对概率统计与随机过程的学习，我不但学习到了非常多的有关概率论的基础知识，也深刻体会到了概率论在生活工作中不可替代的作用。

在我看来，概率论是研究随机现象数量规律的数学学科，而随机现象又是相对于决定性现象而言的。即在基本条件不变的情况下，一系列试验或观察会得到不同的结果。在一系列庞大而又看似不相关的试验数据背后，也许就存在着某些必然的科学本质，而概率论则就是从数据处理入手，以数学的逻辑思维，理性地去揭露观察事物背后的本质。例如，掷一枚硬币是带有偶然性的随机试验，但若多次持续掷一枚硬币，出现正面的频率随着投掷次数的增加逐渐趋于0.5。而大数定律和中心极限定理则是从理论层面上去描述和论证这些普遍性规律的。

随着科学的发展，概率论在生活中的应用越来越广，在生活中发挥越来越广泛的作用，从天气预报，到火箭上天，都离不开概率论。概率论可以用于更好的处理各种不确定因素，给我们生活带来更大的便利，因此，对概率论学科知识的融会贯通，对于我们当代大学生来讲，显得尤为重要。