

计算机与信息学院实验报告

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 实验课程： | 模式识别 | | | | |
| 实验编号： | 6 | | | | |
| 实验名称： | 实验6：参数估计实验 | | | | |
| 实验人员： | 年级 | | 2018级 | | |
| 专业 | | 计算机科学与技术 | | |
| 学号 | | 18111207248 | | |
| 姓名 | | 吴钰 | | |
| 实验日期： | 2021.06.10 | | | | |
| 上交日期： | 2021.6.20 | | | | |
| 实 验 室： | 2060302 | | | | |
| 实验评价： |  | | | | |
|  | 实验成绩： |  | | 评定日期： |  |
|  | 指导教师： | 郑明 | | | |

一、实验目的

使用MATLAB实现概率密度函数的参数估计和非参数估计相关实验

二、实验环境

MATLAB2012

三、实验内容

（1）参数估计（贝叶斯估计）：

1. 以单变量正态分布为例，首先利用MATLAB中的normrnd函数生成总体分布密度服从正态分布*N*（3.4，2.12）的100个样本数据；
2. 假定*p*（*μ*）~*N*（3.6，0.42），利用贝叶斯估计对其进行参数估计，并对估计值与真实值之间的误差进行分析；
3. 在步骤b中误差分析的基础上，以步骤a中的100个样本为基础，继续添加服从正态分布*N*（3.4，2.12）的样本数据，探索贝叶斯估计值和真实值之间的误差以及样本大小之间的关系。

（2）非参数估计（*kN*-近邻估计）：

基于非参数估计公式：,使用normrnd函数，生成均值为0，方差为1，长度为*N*（*N*=1，16，256，10000）的一维正态随机信号数据。画出*kN*和*N*分别为（1，1），（4，16），（16，256）和（100，10000）下的非参数估计结果图，根据结果图分析采样样本数据量的大小与参数估计值精确与否之间的关系。

四、实验设计

根据题意，本文的主要任务是：参数估计实验。其中第1题的解决思路是已知方差，使用贝叶斯估计方法估计均值，根据其公式进行估计，将样本数逐渐增加，观察和判断随着样本数的增加，其结果的变化趋势。

其中第二题的解决思路是，首先生成不同长度的一维正态随机信号数据，再通过Kn-近邻估计算法，进行非参数估计。

五、实验结果

5.1 实验代码：

要求必要时对代码进行注释。

第一题：

T1.m

%产生分布密度服从正态分布N（3.4，2.12）的100个样本数据

data = normrnd(3.4,2.1,[1,100]);

disp(size(data));

x = normrnd(3.6,0.4,[1,100]);

clear;clc;

%样本服从均值U=3.4，标准差V=2.1的正态分布

%U的先验概率p(u)服从U0=3.6，标准差V0=0.4的正态分布

n = 1000;%假设产生100-100000个样本加入

V = 2.1;

U\_temp = 3.6;

V\_temp = 0.4;

X\_m = zeros(1,n);

for i = 1:n

XN = normrnd(3.4,2.1,1,100\*i); %随机生成符合均值U=3.4，标准差V=2.1的正态分布的N个样本

X\_m(1,i) = X\_mean(XN,V,U\_temp,V\_temp);

end

hold on;

plot(1:n,X\_m);

plot(1:n,ones(1,n)\*3.4);

ylabel('均值的估计值');

xlabel('样本数量');

legend('估计值','真实值');

title('贝叶斯估计值和真实值之间的误差以及样本大小之间的关系');

hold off;

X\_mean.m

function res = X\_mean(X,V,U0,V0)

N = length(X);

Mean\_X = sum(X(:))/N; %求样本均值

res = ((N \* V0^2) / (N \* V0^2 + V^2))\*Mean\_X + (V^2 / (N \* V0^2 + V^2))\*U0; %根据公式估计均值

end

第二题：

T2.m

N = [1 16 256 10000]; %样本量N

kN = N.^(0.5); %N的函数kN

x = -4:0.005:4;

for i = 1:length(N)

p = [];

p = Kn(N(i),kN(i)); %调用近邻估计法函数

if i == 1 %对结果进行图形化处理

subplot(2,2,i);plot(x,p);axis([-3,3,0.001,12]);

else

subplot(2,2,i);plot(x,p);axis([-3,3,0.001,1.5]);

end

title(['N = ' num2str(N(i))])

end

Kn.m

function res = Kn(N,Kn)

x = -4:0.005:4;

R = normrnd(0,1,1,N); %随机生成符合0-1正太分布的N个样本

res = zeros(1,length(x));

for i = 1:length(x)

for k = 1:N

temp(k) = abs(x(i) - R(k)); %计算R中样本到x的距离

end

s(1:N) = sort(temp(1:N)); %将距离按从小到大排序

res(i) = (Kn/N) / (s(Kn) \* 2); %根据公式计算p

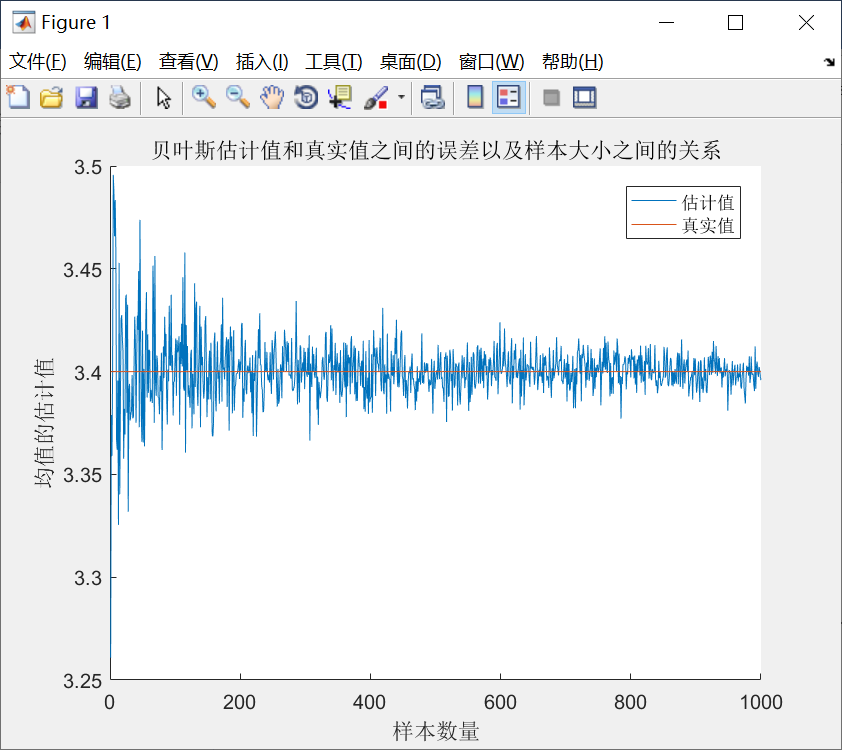
end

end

5.2 结果展示：

实验结果包括输出数据、截图等，将实验文档截图附在此处。

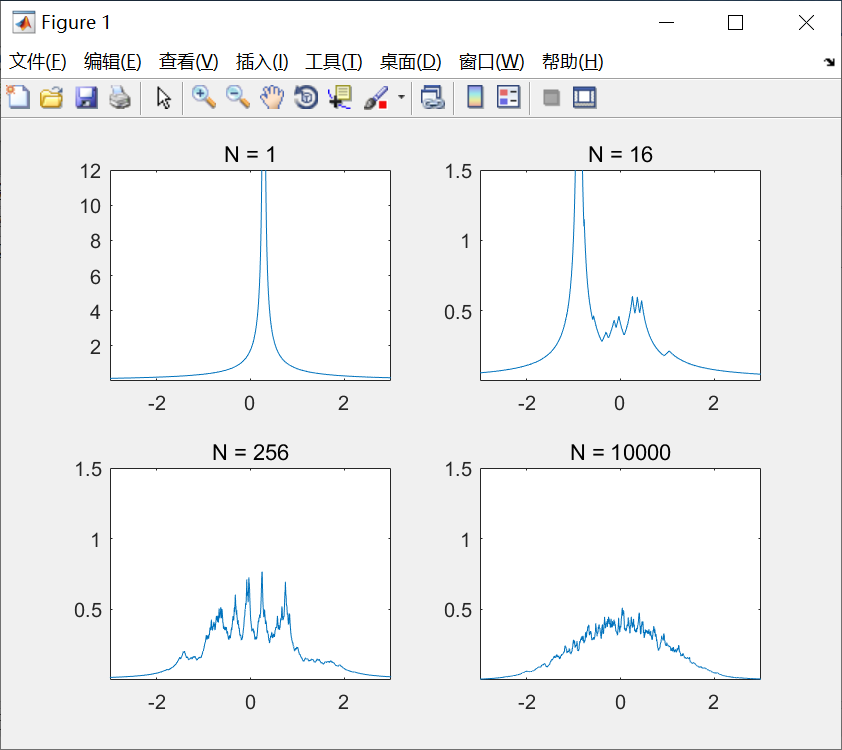
第一题：



结果分析：

随着实验样本数量的增大，最后贝叶斯估计的结果趋向于准确，其上下抖动的幅度减小，在本次实验中为最终的实验均值趋向于3.4。

第二题



结果分析：

通过对于不同样本大小估计出来结果的误差绘图分析，可以明显看出，随着样本的数量增加，其误差值不断减少，在N=1时的误差可以明显可出较大，而在N=10000的计算中，其误差已经较小。

六、实验总结

本次实验主要涉及贝叶斯估计和Kn-近邻估计的操作/内容。

6.1) 两种不同的估计方法，其最终的误差值和其样本数量相关，可以近似的理解为，随着样本数量的增加，其所能得到的误差值不断减小，最终趋向于准确。

6.2) 当需要估计的概率密度函数的形式未知，比如我们并不能知道样本的分布形式时，我们就无法用最大似然估计方法或贝叶斯估计方法来进行参数估计，而应该用非参数估计方法。作为非参数方法的共同问题是对样本数量需求较大，只要样本数目足够大众可以保证收敛于任何复杂的位置密度，但是计算量和存储量都比较大。当样本数很少时，如果能够对密度函数有先验认识，则参数估计能取得更好的估计效果。

6.3) 实验主要参考了书本上P45和P57页的操作，通过理解算法具体步骤来写代码。