CS 215: Data Analysis and Interpretation Assignment: Bayesian Estimation

Report for Problem 1

Deekonda Venkatesh Prasad - Sri Harsha Thota
200050030 - 200050138

Abstract

Finding mean for different estimates
For each n in $N = [5, 10, 20, 40, 60, 80, 100, 500, 10^3, 10^4]$, generate random points from a normal distrubution with $\sigma_{true} = 4$ and $\mu_{true} = 10$ using `randn`(.) function now consider the data generated and assume that the mean is unknown and variance σ_{true} is known to estimate $\hat{\mu}$ for the following priors

1. a Gaussian prior with mean μ_{prior}=10.5 and standard deviation σ_{prior}= 1
2. a uniform prior over $[9.5, 11.5]$

- For each value of N, repeat the experiment 100 times, and plot a box-plot of the error between the true mean μ_{true} and the estimates $\hat{\mu}^{ML}$, $\hat{\mu}^{MAP1}$, $\hat{\mu}^{MAP2}$ where the error is $\frac{|\hat{\mu}-\mu_{true}|}{\mu_{true}}$

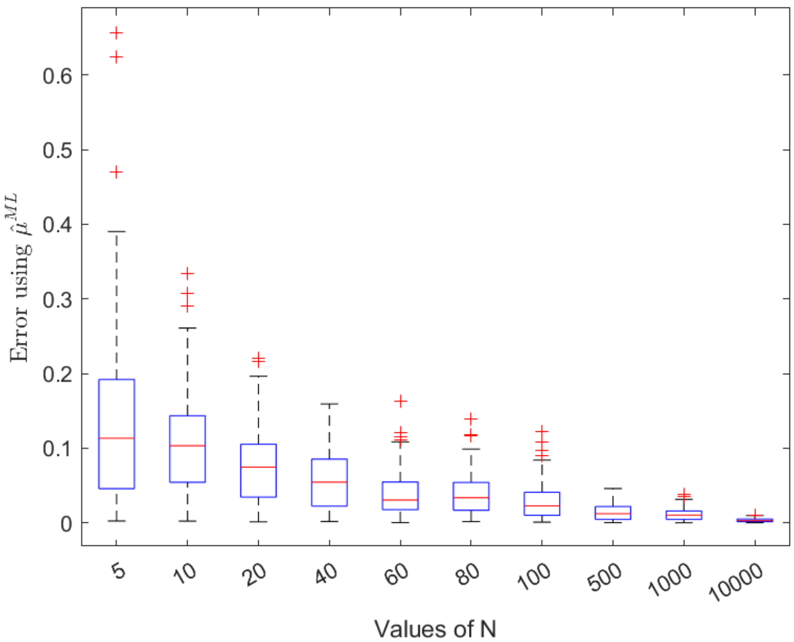
Estimates of μ

$\hat{\mu}^{ML}$ can be simply found out through $\Sigma X_i/N$
$\hat{\mu}^{PAM1}$ can be evaluated using the formula using the formula derived in the class

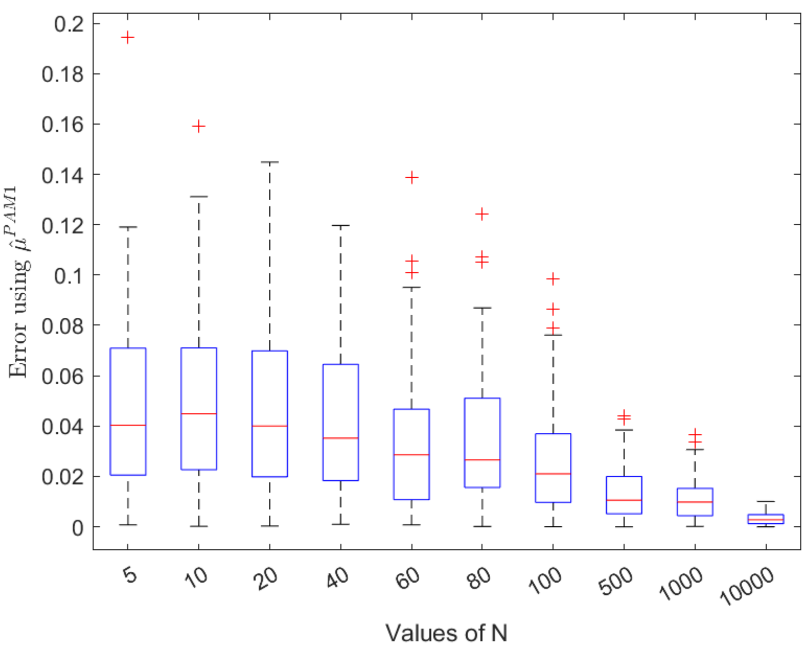
$$\hat{\mu}^{PAM1} = \frac{\bar{x} * \sigma_{prior}^2 + \mu_{prior} * \frac{\sigma_{true}^2}{N}}{\sigma_{prior}^2 + \frac{\sigma_{true}^2}{N}}$$

where $\bar{x}$ is the mean of data generated
and the $\hat{\mu}^{PAM2}$ is same as the maximum likelihood estimate but when the value go outside of the range $[9.5, 11.5]$ we equate it to the closest neighbour
Here are the generated plots showing error vs the value N:[1]

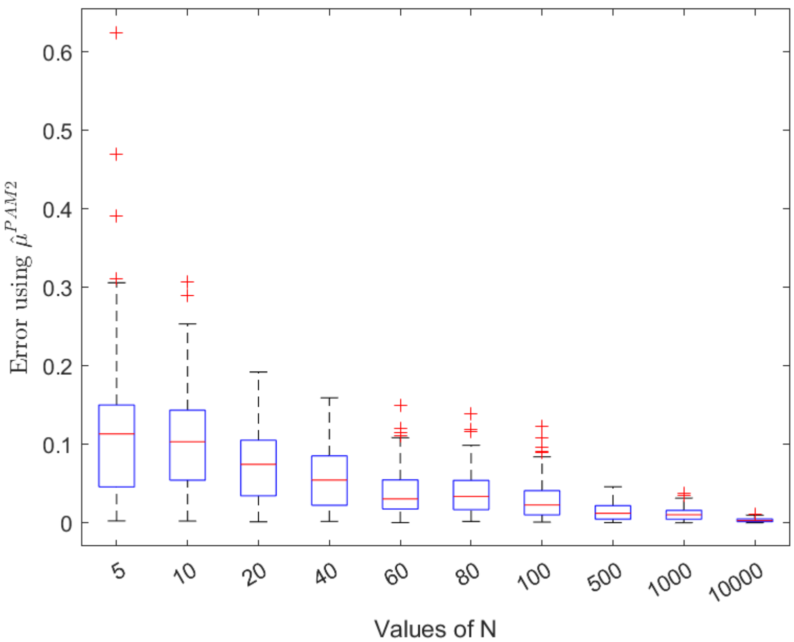
[1] These are also attached in `results` folder as `MLEstimateerror.png, MAP1.png, MAP2.png`



Error using $\hat{\mu}^{ML}$
0.6
0.5
0.4
0.3
0.2
0.1
0
5
10
20
40
60
80
100
500
1000
10000
Values of N



Error using $\hat{\mu}^{PAM1}$
0.2
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
5
10
20
40
60
80
100
500
1000
10000
Values of N



1 Observations

1. As the value of N increases the error almost tends to zero i.e the estimate $\hat{\mu}$ will converge to μ_{true}
2. Of the three estimates i would prefer the $\hat{\mu}^{MAP1}$ because it has less spread than the other two estimates please observe that in the graphs generated the y-axis is of different ranges

Code

Problem2

Please do check the code in folder as the comments are getting cutted off here The following is the MATLAB code for generating uniform random points in ellipse:[2]

```
rng(1);
N = [5,10,20,40,60,80,100,500,10^3,10^4];
sitrue = 4;
muprior=10.5;
sigprior = 1;
mutrue = 10;
```

[2]Attached in the code folder as `q1.m`

```
B = zeros(100,length(N),3);
for i = 1:length(N)
    n = N(i);
    for k = 1:100
        A = 4*randn(n,1)+10;
        muml = sum(A(:,:))/n;
        muap1 = (muml*(sigprior^2)+muprior*(sitrue^2)/n) / (sigprior^2+(si
        muap2 = muml;
        err1 = abs(muml-mutrue)/mutrue;
        err2 = abs(muap1-mutrue)/mutrue;
        if(muap2<9.5)
            mumap2=9.5;
        elseif(muap2>11.5)
            muap2=11.5;
        end
        err3 = abs(muap2-mutrue)/mutrue;
        B(k,i,1) = err1;
        B(k,i,2) = err2;
        B(k,i,3) = err3;
    end
end
fig = figure;
C = reshape(B(:,:,1),100,length(N));
D = reshape(B(:,:,2),100,length(N));
E = reshape(B(:,:,3),100,length(N));
boxplot(C,'Labels',N);
xlabel('Values of N');
ylabel('Error using $$\hat{\mu}^{ML}$$','Interpreter','Latex')
hold on;
saveas(fig,'../results/MLEstimateerror.png');
fig1 = figure;
boxplot(D,'Labels',N);
xlabel('Values of N');
ylabel('Error using $$\hat{\mu}^{PAM1}$$','Interpreter','Latex')
hold on;
saveas(fig1,'../results/MAP1.png');
fig2 = figure;
boxplot(E,'Labels',N);
xlabel('Values of N');
ylabel('Error using $$\hat{\mu}^{PAM2}$$','Interpreter','Latex')
hold on;
saveas(fig2,'../results/MAP2.png');
```