CS 215: Data Analysis and Interpretation Assignment: Bayesian Estimation

Report for Problem 2

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 data x from a uniform distrubution in $[0, 1]$ and now generate a transfomed data sample $y = (-1/\lambda)log(x)$ where $\lambda = 5$

1. find the distribution of the transformed variable y with parameter λ
2. assume a Gamma prior $k(\lambda)$ on the parameter λ with parameters $\alpha = 5.5$, $\beta = 1$

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

Distribution of y

$$y = (-1/\lambda)log(x) \implies x = e^{-\lambda y}$$

this is the $g^{-1}(.)$ function
So the distribution of Y

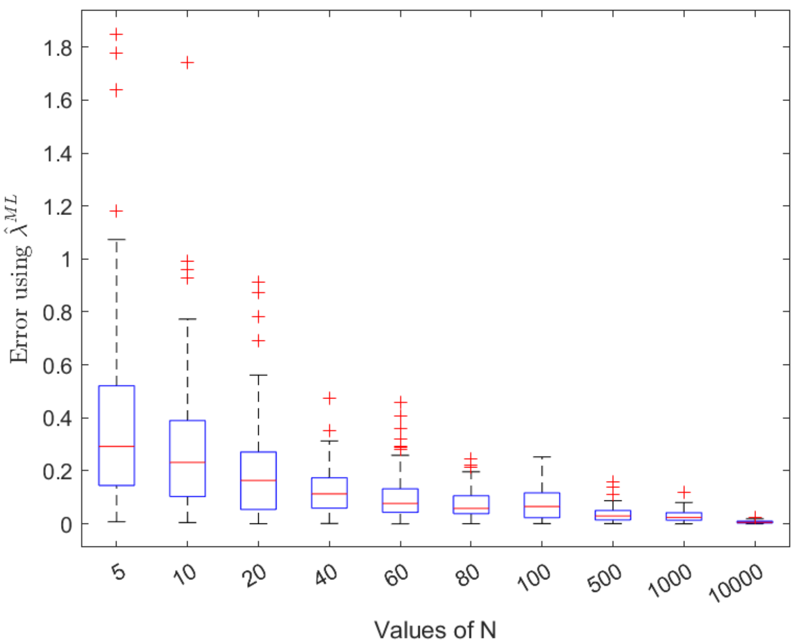
$$q(y) = p(g^{-1}(y))|\frac{d}{dy}g^{-1}(y)| = \lambda e^{-\lambda y}$$

$P(y_1, ...y_n|\lambda) = \prod_{i=1}^{n} q(y_i) = \lambda^n e^{-\lambda \sum_{i=1}^{n} y_i}$

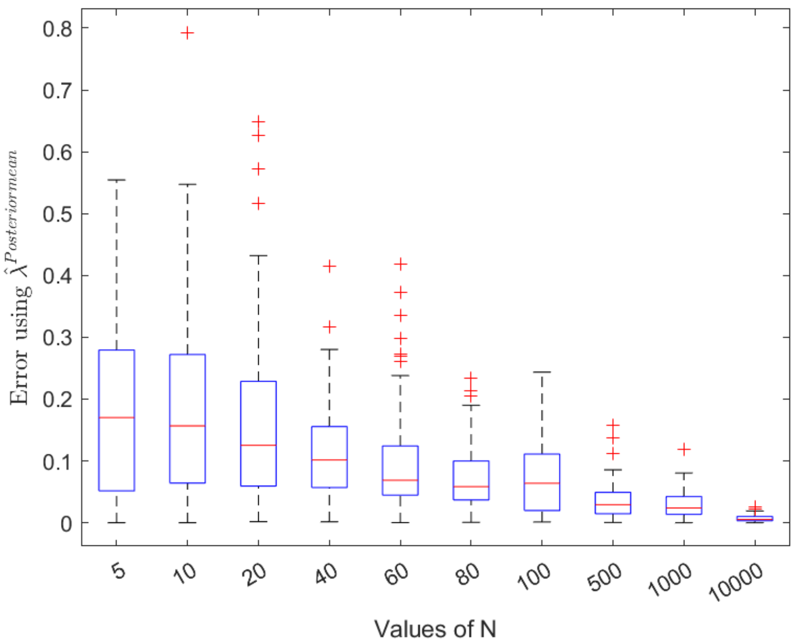
Estimates of λ

the maximum likelihood estimate can be found out by equating $\frac{d}{d\lambda}P(y_1, ...y_n|\lambda)$ to zero and we get $\hat{\lambda}^{ML} = \frac{n}{\sum_{i=1}^{n} y_i}$ for which the double derivative is positive
Finding the posterior distribution
$P(\lambda|y_1, ...y_n) = \frac{P(y_1,....y_n|\lambda)k(\lambda)}{P(y_1,....y_n)}$
here $k(\lambda)$ is the gamma function with parameters $\alpha = 4.5$, $\beta = 1$

The numerator of the posterior distribution is $\frac{\lambda^{n+4.5}e^{-\lambda\sum_{i=1}^{n}y_i+1}}{\Gamma(4.5)}$ and denominator is some constant with parameter n
we get the distribution as $\frac{\lambda^{n+4.5}e^{-\lambda\sum_{i=1}^{n}y_i+1}}{someconstant}$ after sending all the constants to denominator now we can argue that this distribution is a gamma distribution with parameters $\alpha = n + 5.5$, $\beta = \sum_{i=1}^{n} y_i + 1$ as this is already a normalised function so the posterior mean $\hat{\lambda}^{Posteriormean} = \frac{n+5.5}{\sum_{i=1}^{n}y_i+1}$ using the mean formula for gamma function
Here are the generated plots showing error vs the value N:[1]



[1]These are also attached in `results` folder as `ErrorusingMLE.png`, `ErrorUsingPosteriorMean.png`



Observations

1. As the value of N increases the error almost tends to zero i.e the estimate $\hat{\lambda}$ will converge to λ

2. Of the two estimates i would prefer the $\hat{\lambda}^{posteriormean}$ because it has less spread than the $\hat{\lambda}^{ML}$ 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(0)
N = [5,10,20,40,60,80,100,500,10^3,10^4];
lamtrue = 5;
B = zeros(100,length(N),2);
for i=1:length(N)
    n = N(i);
```

[2]Attached in the code folder as `q2.m`

```
    for m = 1:100
        x = rand(n,1);
        y = (-1/5).*(log(x));
        lamml = n/sum(y(:));
        lampam = (n+5.5)/(sum(y(:))+1);
        err1 = abs(lamml-lamtrue)/lamtrue;
        err2 = abs(lampam-lamtrue)/lamtrue;
        B(m,i,1) = err1;
        B(m,i,2) = err2;
    end
end
C = reshape(B(:,:,1),100,length(N));
D = reshape(B(:,:,2),100,length(N));
fig = figure;
boxplot(C,'Labels',N);
xlabel('Values of N');
ylabel('Error using $$\hat{\lambda}^{ML}$$','Interpreter','Latex')
hold on;
saveas(fig,'../results/ErrorusingMLE.png');
fig1 = figure;
boxplot(D,'Labels',N);
xlabel('Values of N');
ylabel('Error using $$\hat{\lambda}^{Posteriormean}$$','Interpreter','Late
hold on;
saveas(fig1,'../results/ErrorUsingPosteriorMean.png');
```