Analyzing loan information and night light intensities in Nigeria: A Bayesian Approach

Bayesian Statistics Tutorial

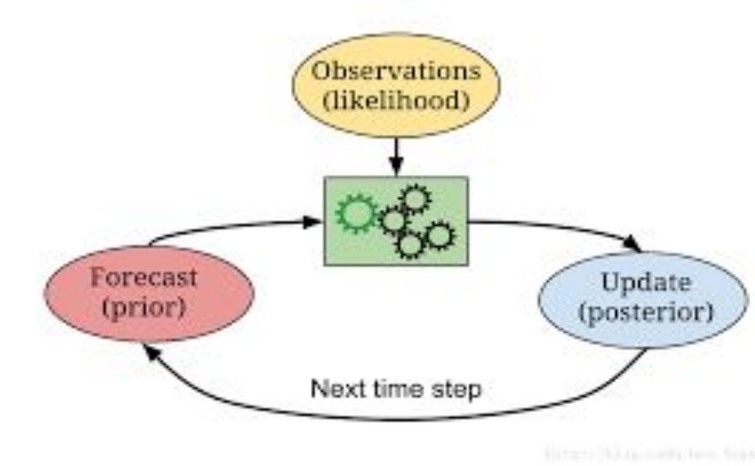
Aayush Baral

Introduction

Artificial Night Light intensities are a good indicator of prosperity in the developing world. Places where it is difficult to do a survey, these metrics can be used to predict the inequality and poverty level. Kiva, which is a micro funding platform that lends money via the internet to low-income entrepreneurs and students in over 80 countries. Analysis of the relationship between nightlight data and economic indicators of various areas will help Kiva to detect places where there is more need for flow of loans. This project aims to look if there is any significant difference in average loan provided to places with light intensities threshold of three using Bayesian Inference for difference between the means.

Bayesian vs Frequentist Approach

The distinction between Bayesian and Frequentist is associated with the definition of probability. For frequentists, probabilities are fundamentally related to the frequency of the occurrence of events whereas, for Bayesians the definition of probability is extended to cover the degrees of uncertainty about statements. Frequentist consider model parameters to be fixed but data to be random, while Bayesian consider the data to be fixed and parameters to be random. Bayesian approach starts with a prior knowledge about the probability distribution of an event.



Data and Methodology

The data contains different clusters in Nigeria with household level information and was obtained from Kaggle. We divide the dataset into two samples based on our nightlight composite threshold of 3 and assume that they have unequal and unknown variances. We compare the mean of these two sample distributions. We use the complete dataset to build the Bayesian Regression Model to see the relationship between the nightlight composites and the amount of loan lended.

Analysis

We conducted a two sample Bayesian t-test for our two samples of data using uniform prior. As both of our samples were quite large, the t-distribution can be represented as a z-distribution.

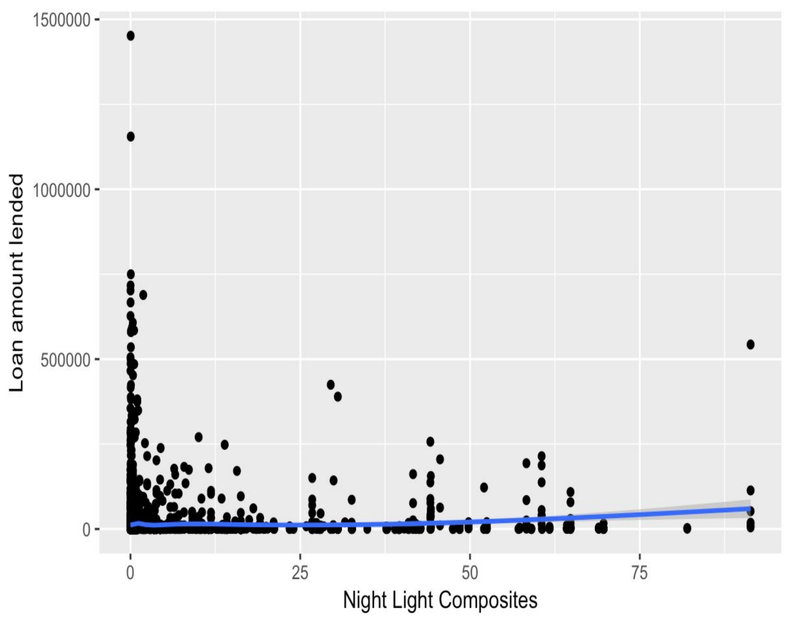
From the analysis, we found out that there is not a huge difference between our posterior mean of the samples. Hence, the whole distribution can be represented using a linear function. Hence, we build a Bayesian linear model to look at the relationship between nightlight composites and loan amount.

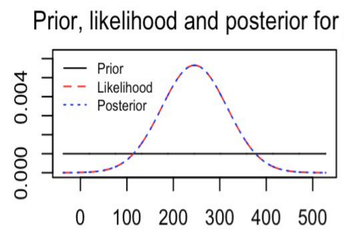
Results and Conclusion

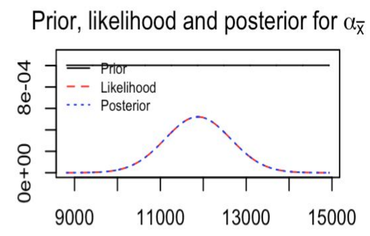
From our Bayesian t-test, we found that the difference between the mean values of each sample is not zero. Further, we build a Bayesian regression model which showed the causal relationship between the night light composites and the loan amount. The model gave us a 95% credible interval which explains that there is a 95% probability that our prediction falls within out CI bounds.

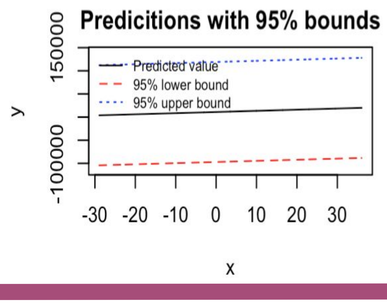
```
        Two Sample t-test

data:  updated_data1$amount and updated_data2$amount
t = -2.5245, df = 5195, p-value = 0.01162
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -9899.066 -1245.052
sample estimates:
posterior mean of x posterior mean of y
           11096.64            16668.70
```

```
Standard deviation of residuals:  55100
           Posterior Mean Posterior Std. Deviation
           -------------- ------------------------
Intercept:  11880          764.75
Slope:      245.4          70.629
```









References
https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics-spring-2014/readings/MIT18_05S14_Reading20.pdf