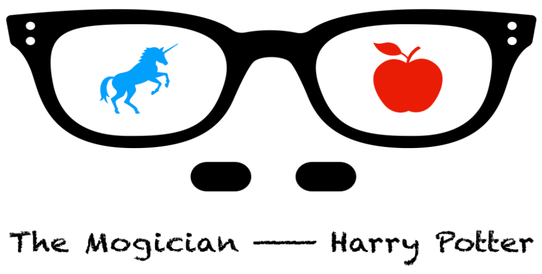
# VE414 • Bayesian Analysis • Project

**Author:** Wang Xuanyu, Wu Guangzheng, Jiang Sifan
**Instructor:** Jing Liu




## Abstract

In this project, we are using the track data to predict the position of trees. We are given 43 trips with about 500 points per trip as the input data. Data preprocessing is done in order to deal with the overlap of the data points in one trip. Bayesian model is applied to predict the position and number of hidden trees.

## Key Word

Prediction with track data, Bayesian Model

## Problem Statement

Hidden trees are required to be estimated in a garden. Travellers can detect the number of fruits, which is produced by the hidden tree, near the traveller, at some certain position, and estimate the number of trees in the garden.

## Data Visualization

We use matplotlib to plot the data points from one trip in the figure. During each trip, the traveller only went through a certain part of the region, near the $y = x$ line. Regions showing the "nearest" result of different data points may also overlap with one another.

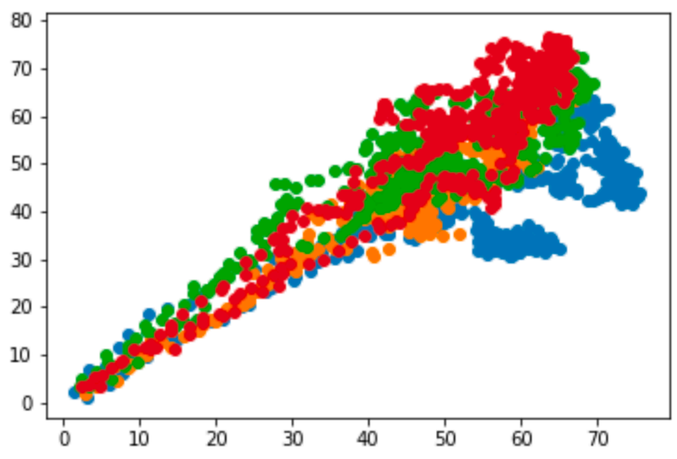



**Fig. 1:** Plot of Data Points in Trip 1-4

Evidence can be provided to prove that the number and location of fruits didn't change in the whole process. That is, no extra fruit is produced by the hidden trees, and the travellers do not pick away the existing fruits.

## Data Preprocessing

Data Preprocessing is done in order to reduce the impact of errors and noise from the data. From the previous sections, we found that data points overlap with one another, making the fruits double-counted by different data points. Without data preprocessing, the estimated number of trees will be larger.

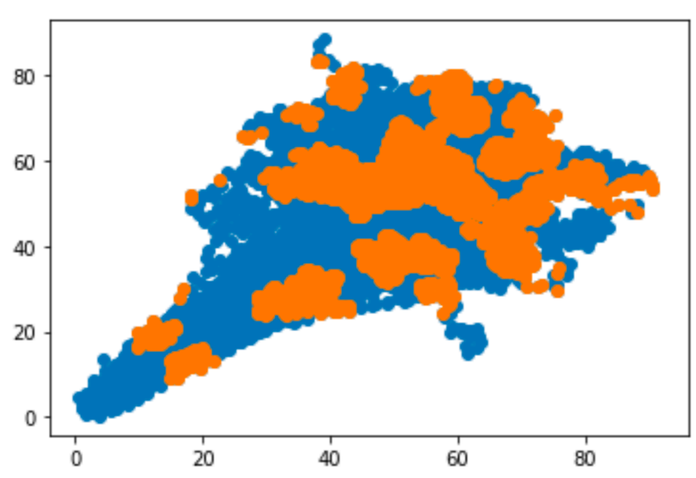



**Fig. 2:** Data points overlap with each other

We reduced the number of data points by deleting some data points to avoid overlapping. Whenever two "close" regions overlap, the data points showing less fruit will be deleted, such that no space in the graden will be counted by two different data points.

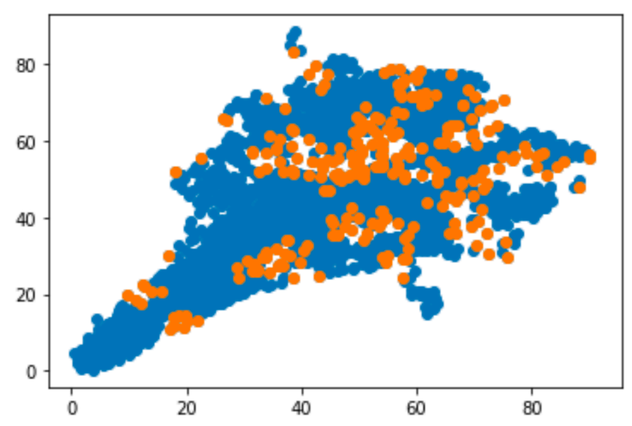



**Fig. 3:** Reduced Data Points without Overlap

## Modeling and Analysis

We use a Bayesian model to predict the number of trees. Suppose for every point in the whole garden, there is a probability density $p_k$ here that at this point a tree exist, such that the reject probability should be $1 - p_k$ here.

Since we had avoided the overlap, so we can imagine that for each data point, the place of the fruit follows a uniform distribution, and each fruit can be considered as a likelihood to update the posterior of $p_k$. A fruit will increase the probability of nearby $p_k$, while an empty data point will reject the probability $p_k$.

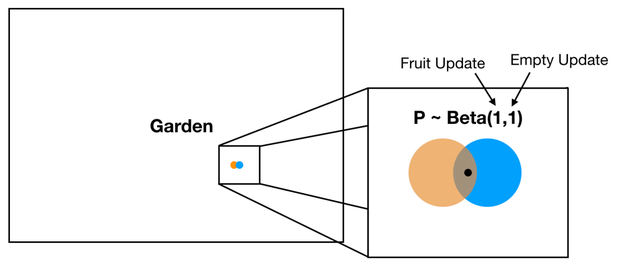



**Fig. 4:** Our Bayesian Model

The following table shows some parts of the resulted mean after an update of data points from the first trip. We can see that the points without being met remains unchanged, while probability density at point met the empty point was reduced.

| Point | 2 | 3 | 4 |
|---|---|---|---|
| Position | $(0, 2)$ | $(0, 3)$ | $(0, 4)$ |
| Fruit | 0 | 0 | 0 |
| Empty | 0 | 0 | 1 |
| $\alpha$ | 1 | 1 | 1 |
| $\beta$ | 1 | 1 | 2 |
| mean | 0.500 | 0.500 | 0.333 |

**Table 1:** $p_k$ after a First Update

After the final update, we get a set of probability $p_k$ to as the sample of probability density. Each $p_k$ identifies the probability that at that point a tree is existed. The total number of trees should be the integral of the $p_k$ over the garden. We estimated that there are 169 trees that has been visited.

The region visited is about $1/3$ to $1/2$ of the whole garden, which is shown in figure 2. So we estimated that there are about 400 trees in the garden.

## Conclusion

This model provides a bayesian approach to predict the number of trees, which can also be applied to the situation where trees can move. Further improvement should be made to find a better way to deal with the overlap, since it is too strong and direct now.