VE414 Group Project

VE 414 Group3

Dec.5th , 2019

Xia Peiyan, Hu Bingcheng, Zhao Xuanyi

JOINT INSTITUTE
交大密西根学院

0

PART ZERO

Outline

0

PART ZERO

Outline

- △ Problem Overview
- ▲ Data set
- △ Early Estimation
- ▲ GMM & EM method
- △ Poisson method
- ▲ Further discussion

1

PART ONE

Problem Overview

Invisible trees & visible fruits

Spells that records fruits’ position

Multiple travels

Estimate the number of trees

1

PART ONE

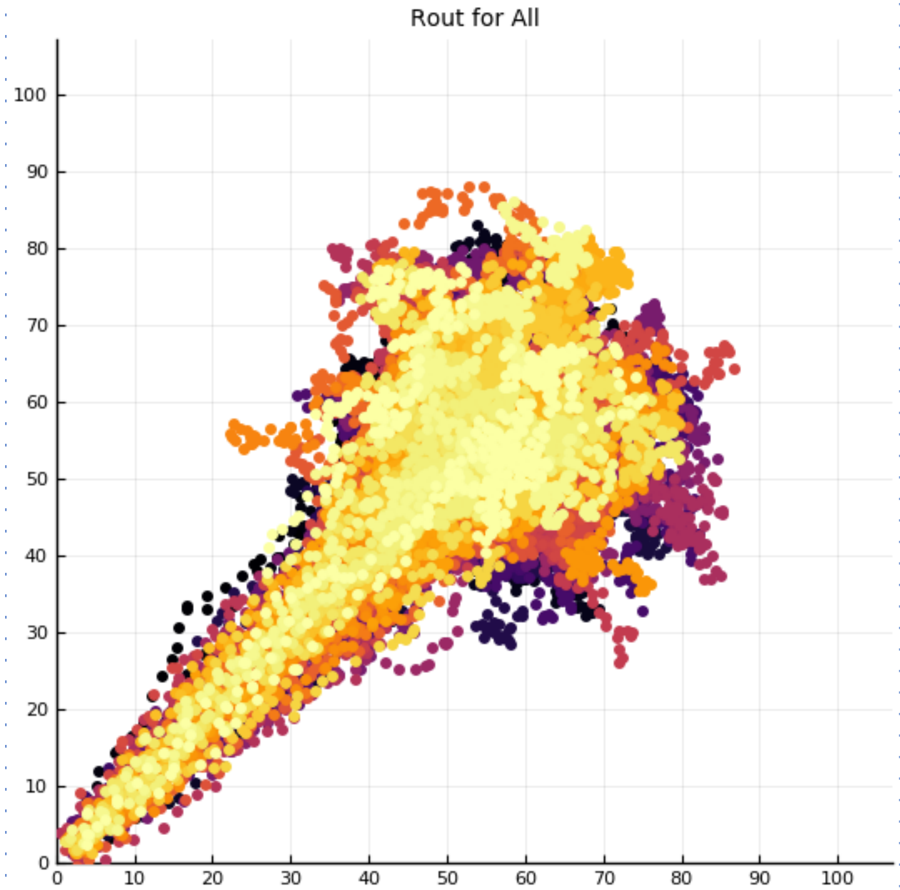
JOINT INSTITUTE

交大密西根学院

Outline

- △ Problem Overview
- ▲ Data set
- △ Early Estimation
- ▲ GMM & EM method
- △ Poisson method
- ▲ Further discussion

Data Set



Basic Info:

- 107*107 grid
- Numbers of fruits within 1 meter
- Numbers of fruits within 3 meters
- Multiple travels

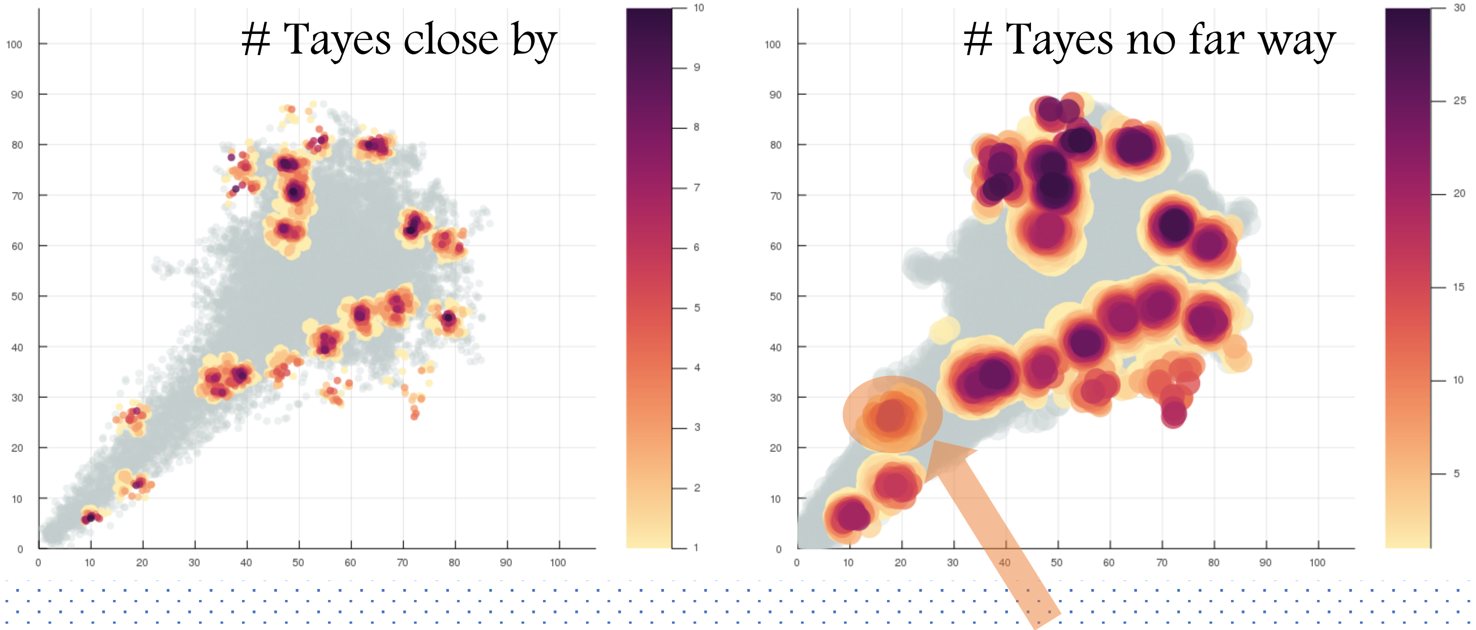
2

PART TWO

Data Set

Graphs of Tayes numbers:

- Gray area means no Tayes
- Yellow area indicates at least one Taye
- Purple area means a lot of Tayes
- Intuitively, there should be at least one Jiuling at the center of each cluster



One cluster

2

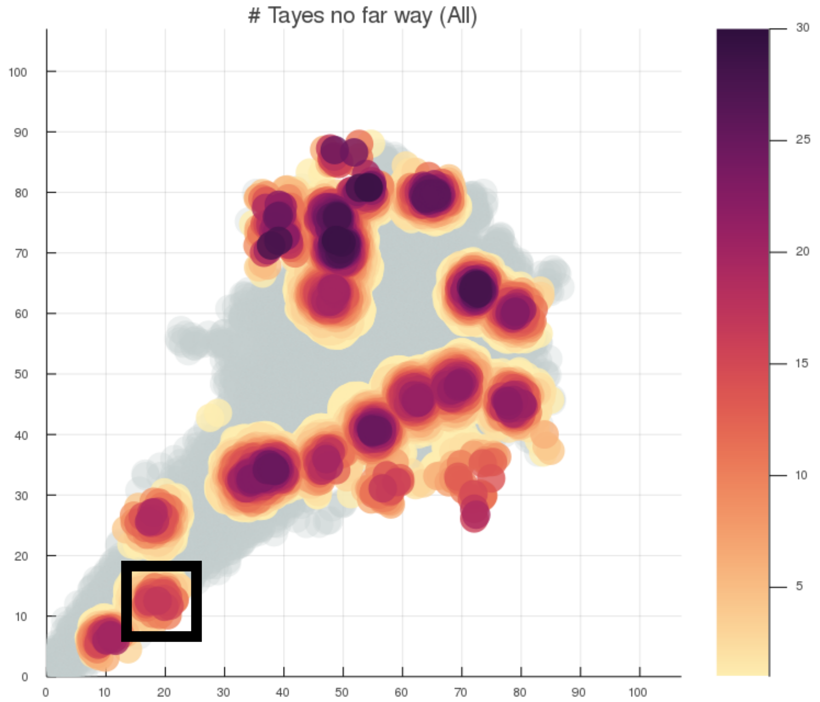
PART TWO

Outline

Early Estimation

Assumptions:

- Take one area as a sample for a single tree
- The block with most fruits observed is the tree's position
- The average number of fruits in 1x1 area can be get by calculating the (observed fruits within far distance/observation times)

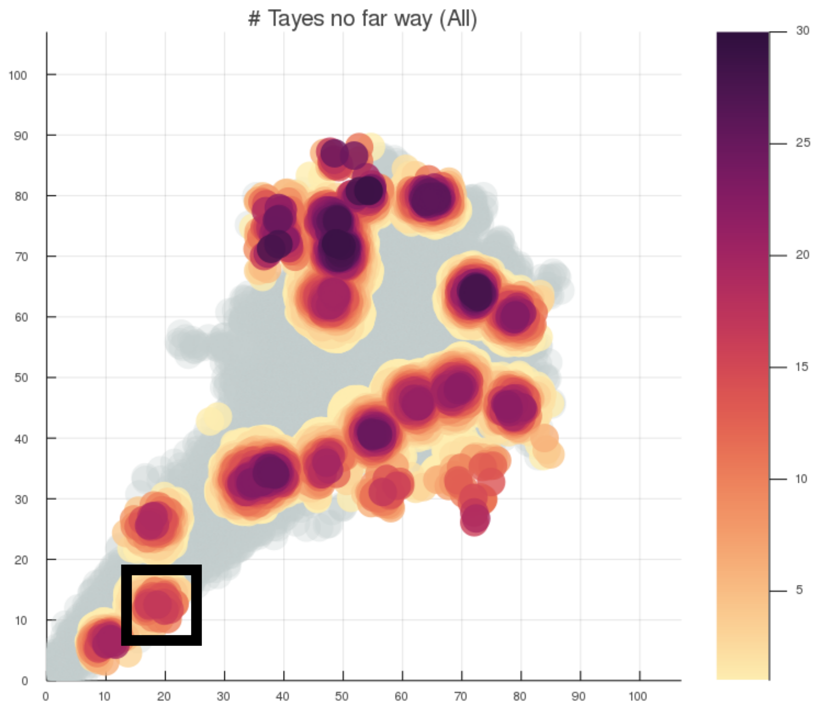


Early Estimation

Estimation:

- $\frac{Area\ with\ friuts(far)}{Area\ of\ sampled\ tree} * \frac{fruit\ per\ area\ in\ sampled\ tree}{fruit\ per\ area\ in\ (far>0)}$

- Rough estimation: 21 trees



3

PART THREE

Outline

- △ Problem Overview
- ▲ Data set
- △ Early Estimation
- ▲ GMM & EM method
- △ Poisson method
- ▲ Further discussion

GMM

JOINT INSTITUTE
交大密西根学院

Intro:

- a type of clustering algorithm
- each cluster is modelled according to a different Gaussian distribution
- soft assignments rather than hard assignments (k-means)

GMM

- GMM Alogorithm

- Introduce a latent variable γ(gamma) for each data point

- It assumes that each data point was generated by some information about γ(gamma). It tells us which Gaussian generated a particular data point.

GMM

Goal:
Maximise the marginal likelihood of data points X given our params.

Consider log likelihood of Linear super-position of Gaussians

$$ln\, p(X|\mu, \Sigma, \pi) = \Sigma_{n=1}^{N} ln \left\{ \Sigma_{k=1}^{K} \pi_k N(x_n|\mu_k, \Sigma_k) \right\}$$

EM

However, in practice, we do not observe these latent variables so we need to estimate them.

Expectation Maximisation (EM) Algorithm

EM

JOINT INSTITUTE
交大密西根学院

E~STEP:

Recall that

$$p(x) = \Sigma_{k=1}^{K} \pi_k N(x \mid \mu_k, \Sigma_k)$$

Therefore:

$$\gamma(z_{nk}) = \frac{\pi_k N(x_n|\mu_k, \Sigma_k)}{\Sigma_{j=1}^{K} \pi_j N(x_n|\mu_j, \Sigma_j)}$$

EM

M-STEP:

Loop the steps until convergence.

$$\mu_k^{new} = \frac{1}{N_k} \Sigma_{n=1}^{N} \gamma(z_{nk}) x_n$$

Mean of the Gaussians

4

PART FOUR

EM

M-STEP:

Loop the steps until convergence.

$$\Sigma_k^{new} = \frac{1}{N_k} \Sigma_{n=1}^{N} \gamma(z_{nk})(x_n - \mu_k^{new})(x_n - \mu_k^{new})^T$$

Covariance of the Gaussians

EM

JOINT INSTITUTE
交大密西根学院

M-STEP:

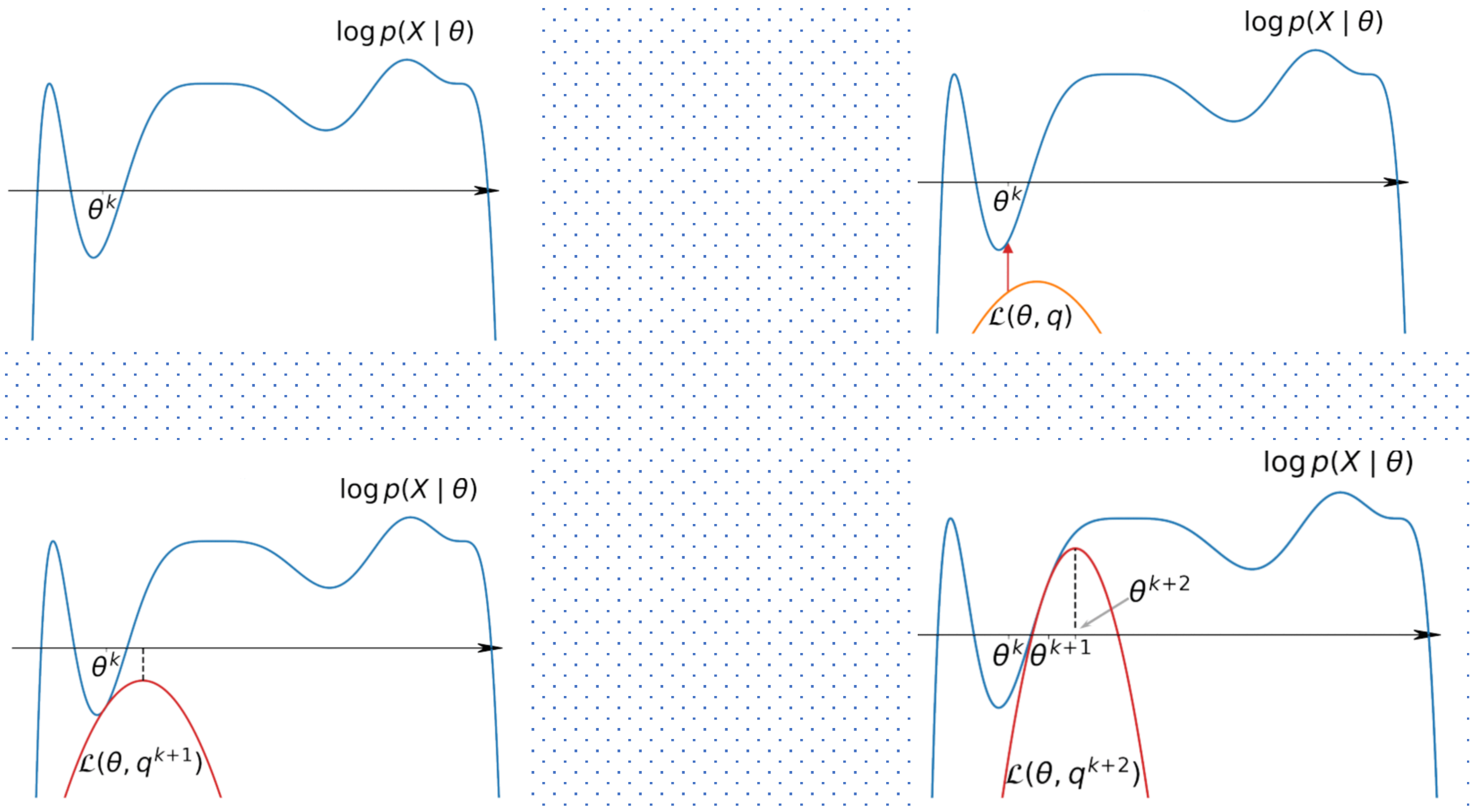
Loop the steps until convergence.

$$\pi_k^{new} = \frac{N_k}{N}$$

$$N_k = \Sigma_{n=1}^{N} \gamma(z_{nk})$$

weights and sum of responsibilities in each Gaussian k

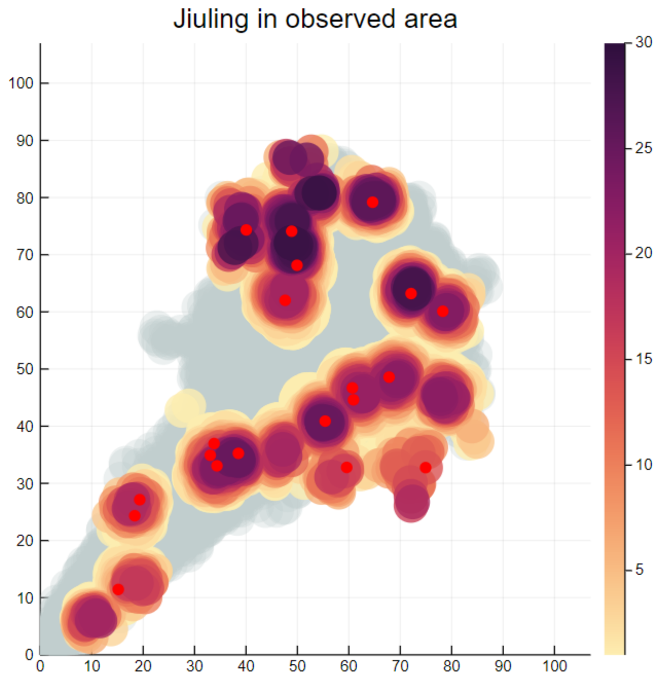
GMM+EM


$\log p(X \mid \theta)$
θ^k
$\log p(X \mid \theta)$
θ^k
$\mathcal{L}(\theta, q)$
$\log p(X \mid \theta)$
θ^k
$\mathcal{L}(\theta, q^{k+1})$
$\log p(X \mid \theta)$
θ^{k+2}
θ^k θ^{k+1}
$\mathcal{L}(\theta, q^{k+2})$

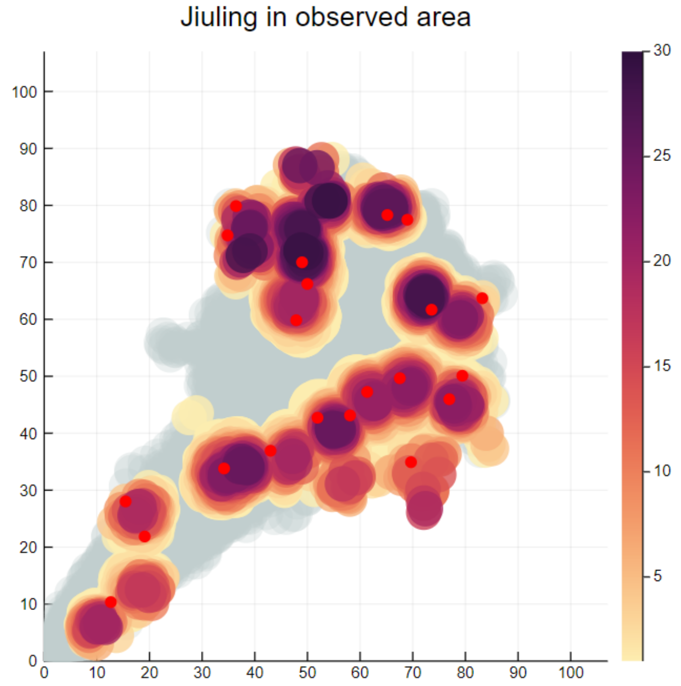
4

PART FOUR

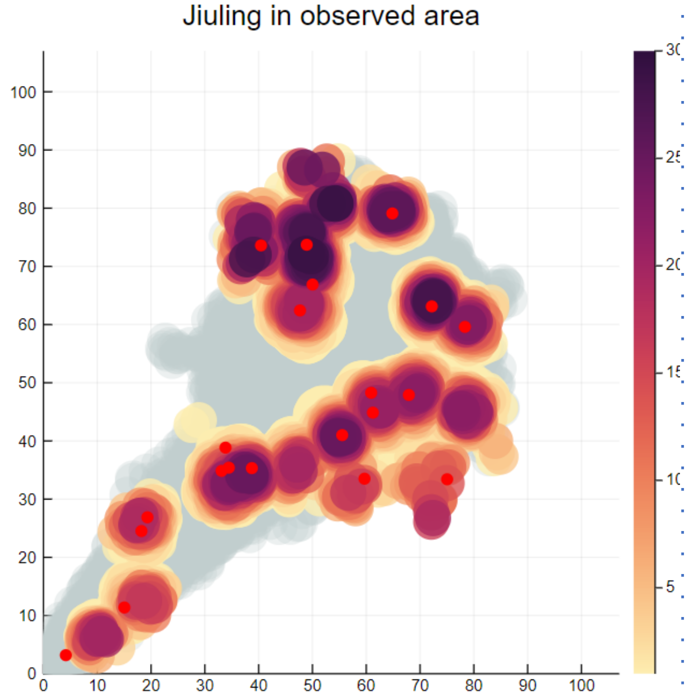
Results



Consider only near data



Consider only far data



Consider weighted sum of near and far data

4

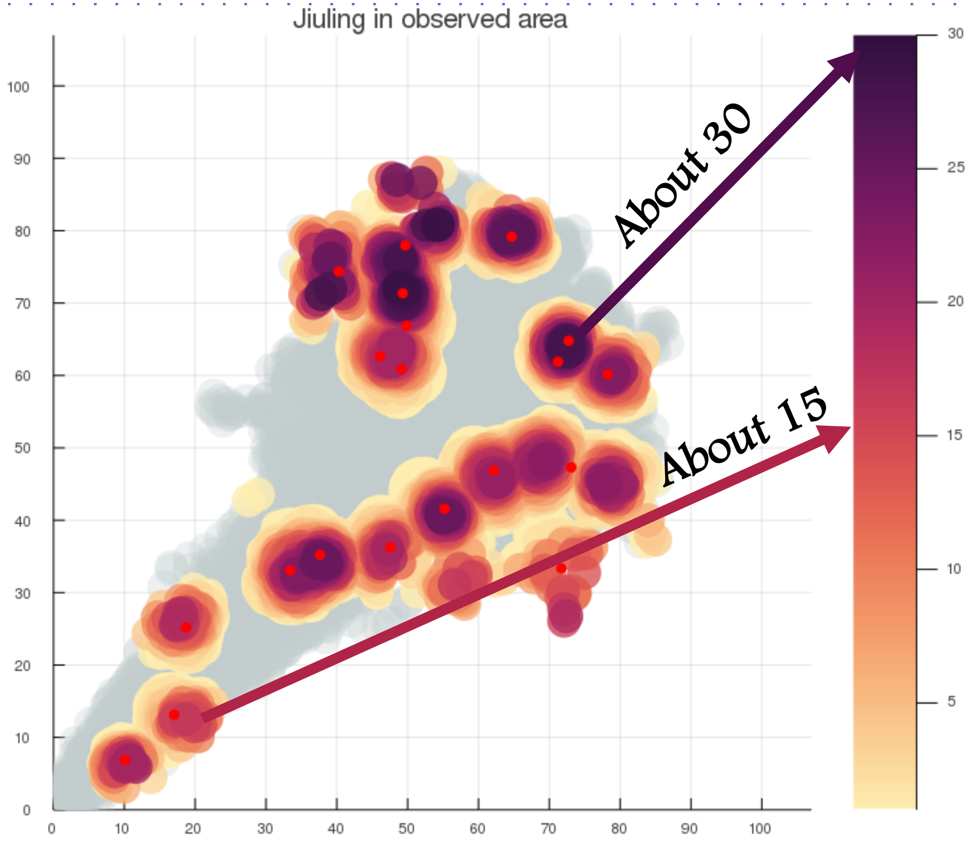
PART FOUR

Outline

JOINT INSTITUTE
交大密西根学院

Assumptions

- Two Tayes can live at one position
  - The center of clusters have different #Tayes
  - It's reasonable to consider there are multiple Jiulings
- The probability of the existence of Jiuling is the same inside and outside the observed area
- There can not be any Jiuling on the gray area



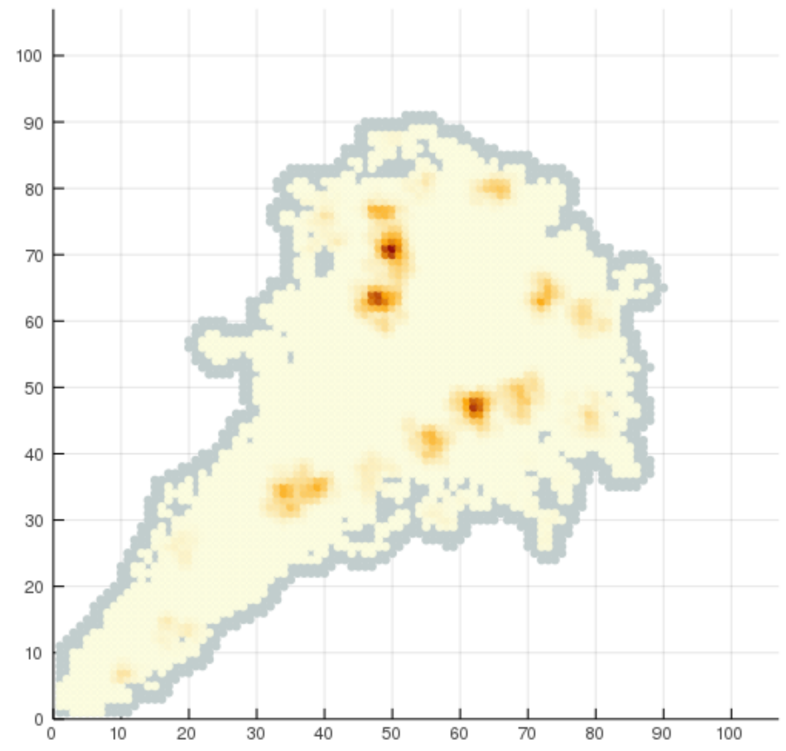
Estimate Grid

Consider the following model of the unobserved area:

$$X_i \sim Poisson(\lambda)$$

Where λ is the expectation of a conjugate prior:

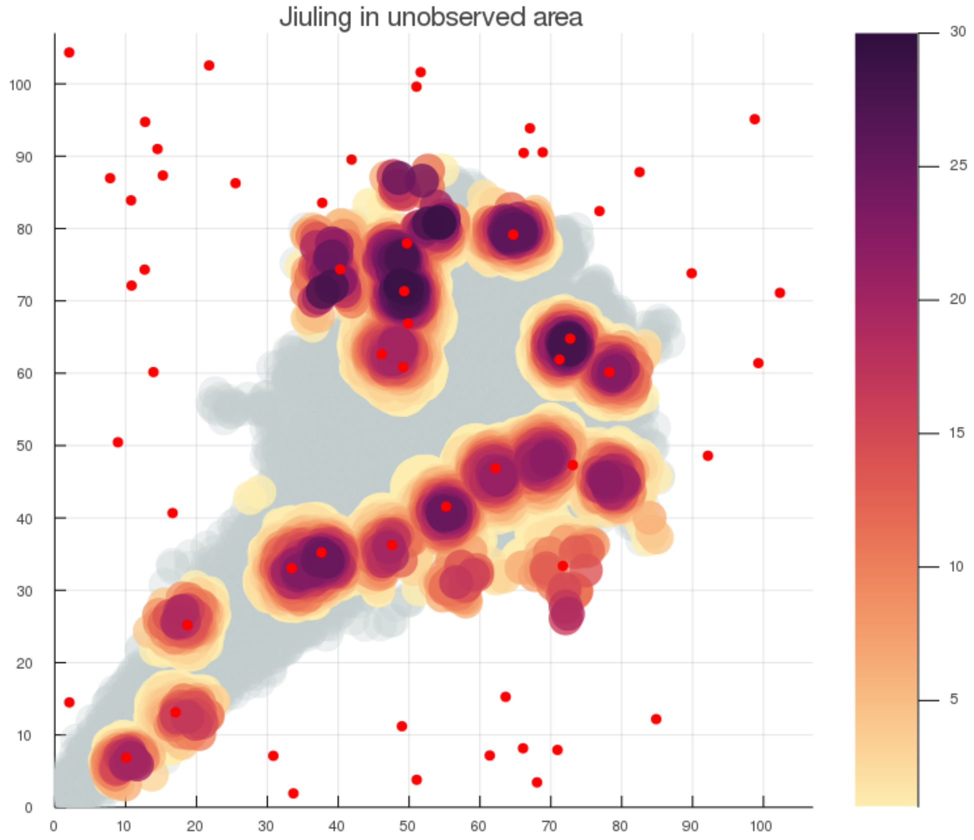
$$\lambda \sim E(Gamma(\alpha, \beta))$$



Random Jiulings

With the estimated Poisson distribution, we estimate the number of trees outside of the observed area

Every turn, the positions are different, but the number of Jiulings are similar



21 observed +37 unobserved

5

PART FIVE

Outline

JOINT INSTITUTE
交大密西根学院

Walking Jiuling

What if Jiuling can actually move ?

Just like “Socratea”, which can “walk” 20 meters a year!

Img.baidu.com/KJFSN.98SFU8D

6

PART SIX

JOINT INSTITUTE
交大密西根学院

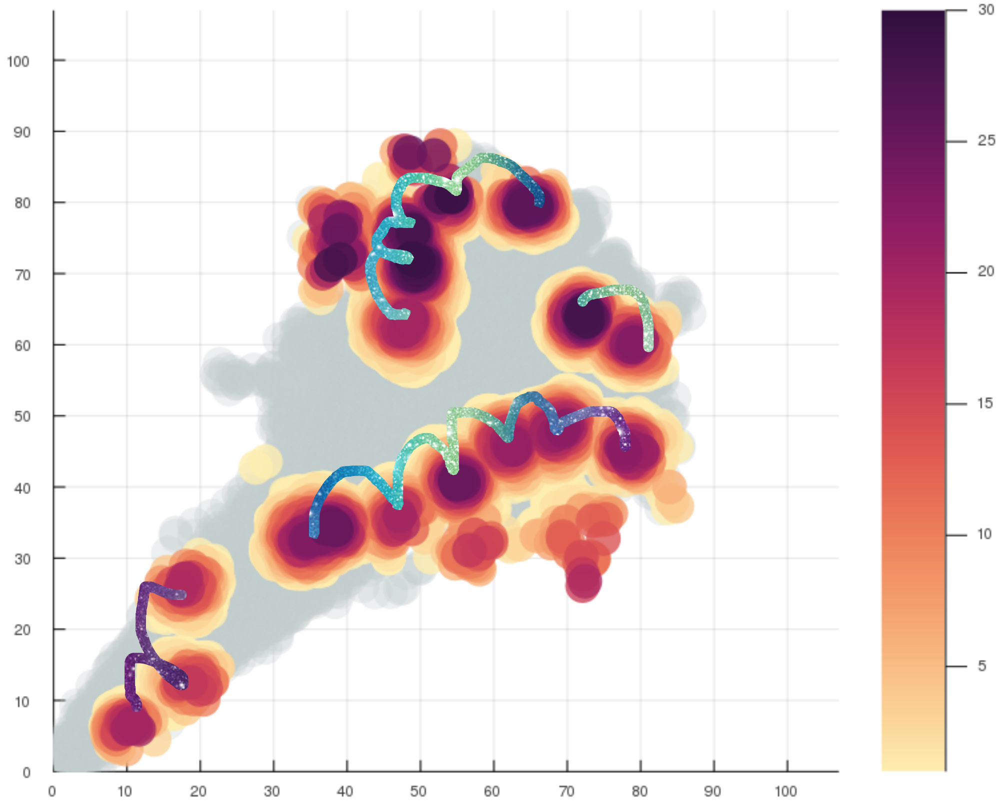
Walking Jiulings

~ Do Jiulings walk or jump?

If they walk, each term there position should be different, so the finding of Tayes should not be clusters

~ Do they choose direction randomly?

~ Is there speed fast or slow?



7

PART SEVEN

JOINT INSTITUTE
交大密西根学院

Q&A

Reference

https://towardsdatascience.com/gaussian-mixture-modelling-gmm-833c88587c7f