DEVELOPING AN EXCLUSION BOUNDARY (EB) FOR THE DASHER-XLS

STAT 938

Alaattin Isilak, Kyu Min Shim, Pamalka Bandara, Sohoon Youn

Team Exclusion Boundary Hunters

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

Outline

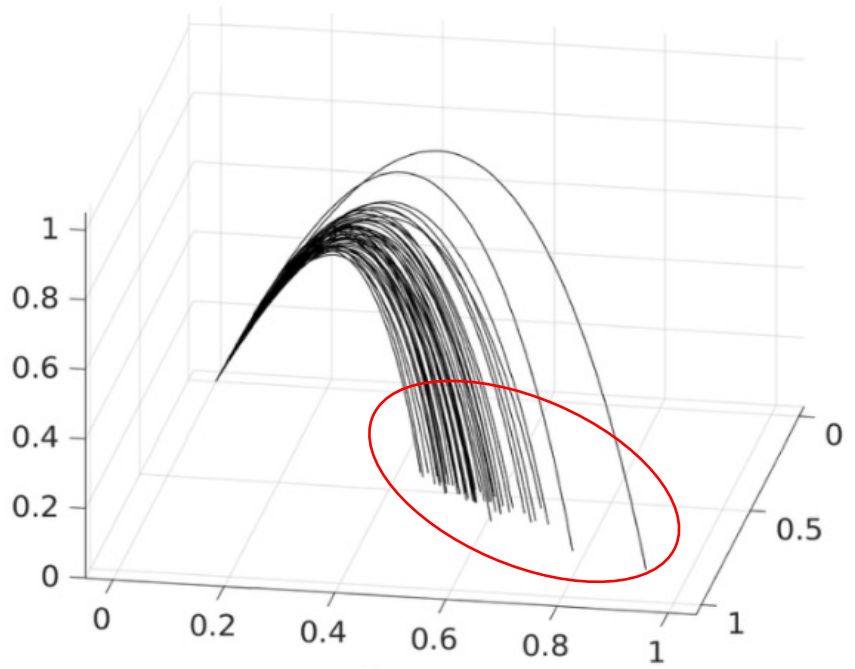
1. Introduction
2. Description Problem and Data
3. Methodology
4. Results
5. Conclusions

1. Introduction

2. Description of Problem

- **Problem**

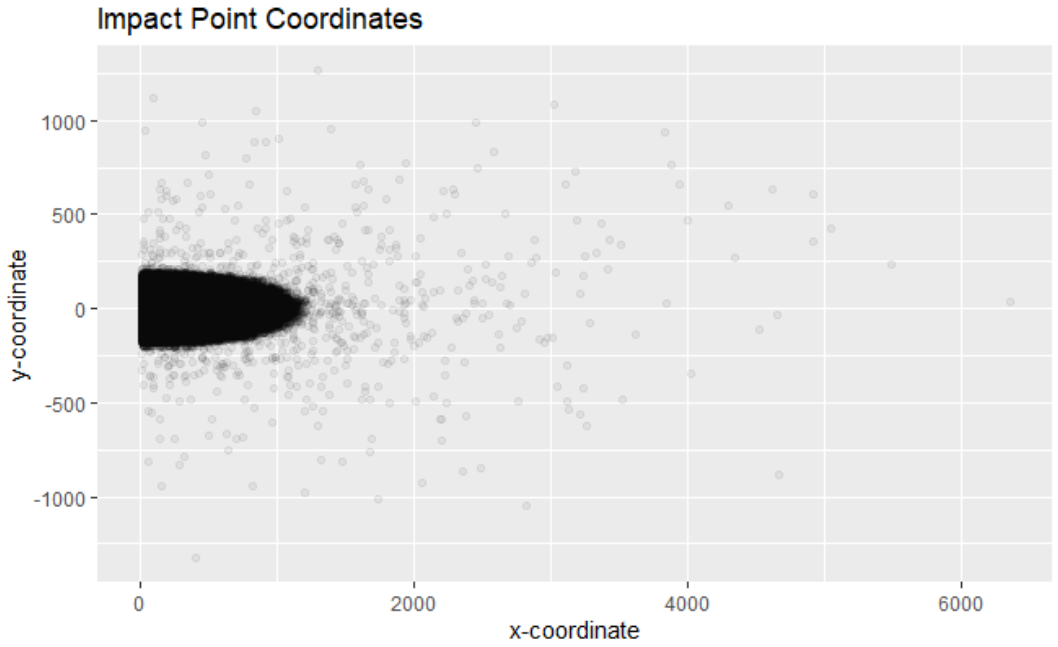
Define an exclusion boundary with a confidence level of γ% and a probability of crossing θ based on the impact point data



2. Description of Data

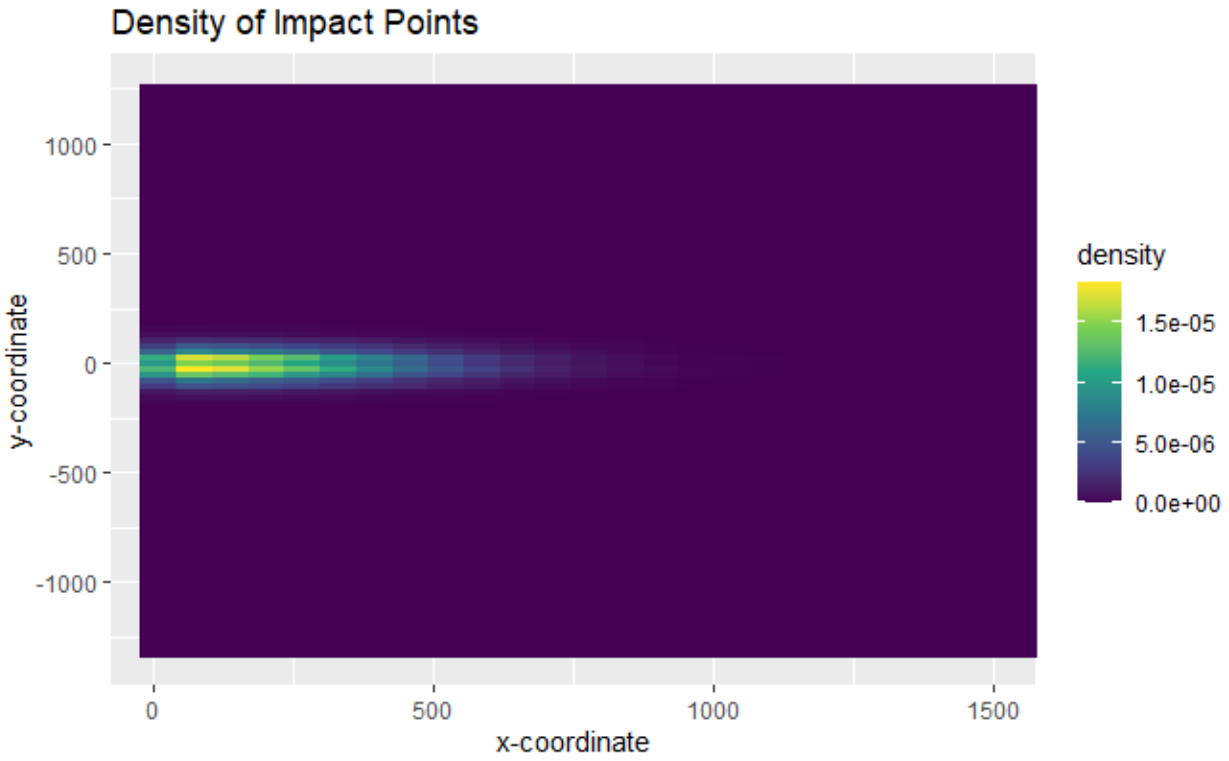
- **Data**

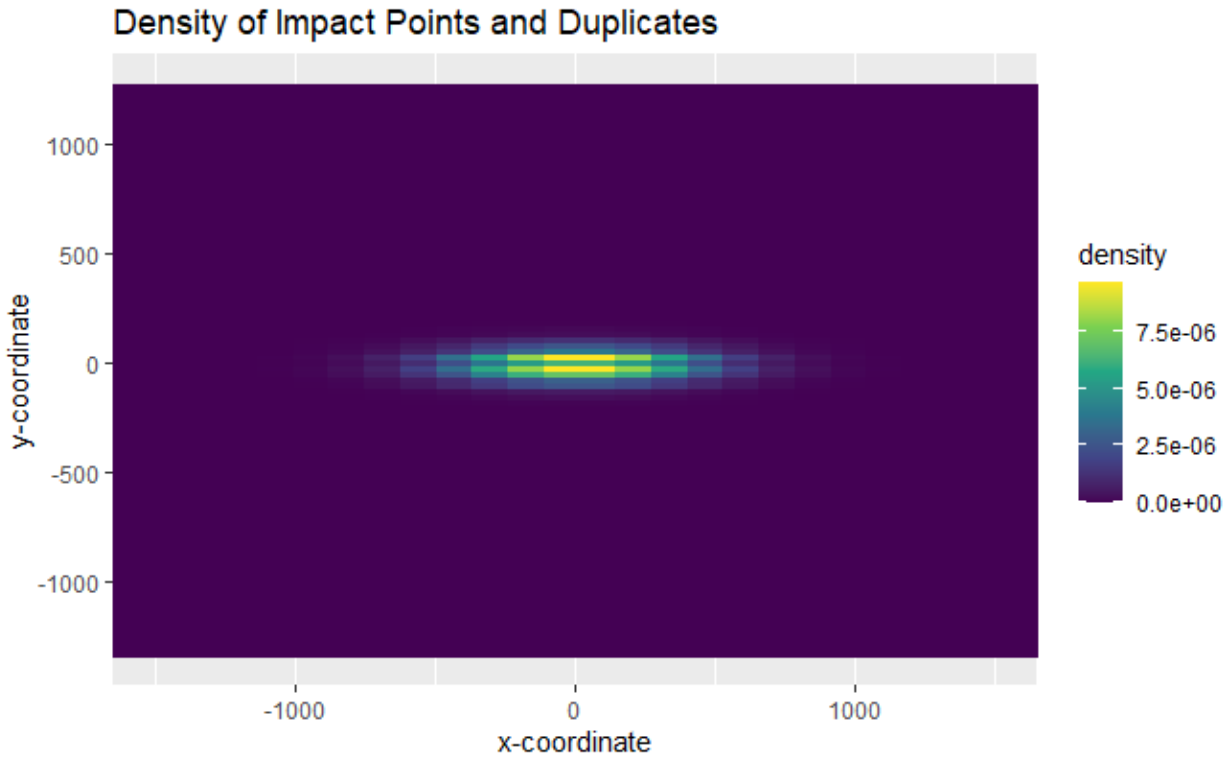
The impact point coordinates (x,y) for multiple Dasher-XLS launches



	X	Y
Min	5	-1325
Max	6355	1265
Median	215	5
Mean	251.7	-0.01

2. Description of Data





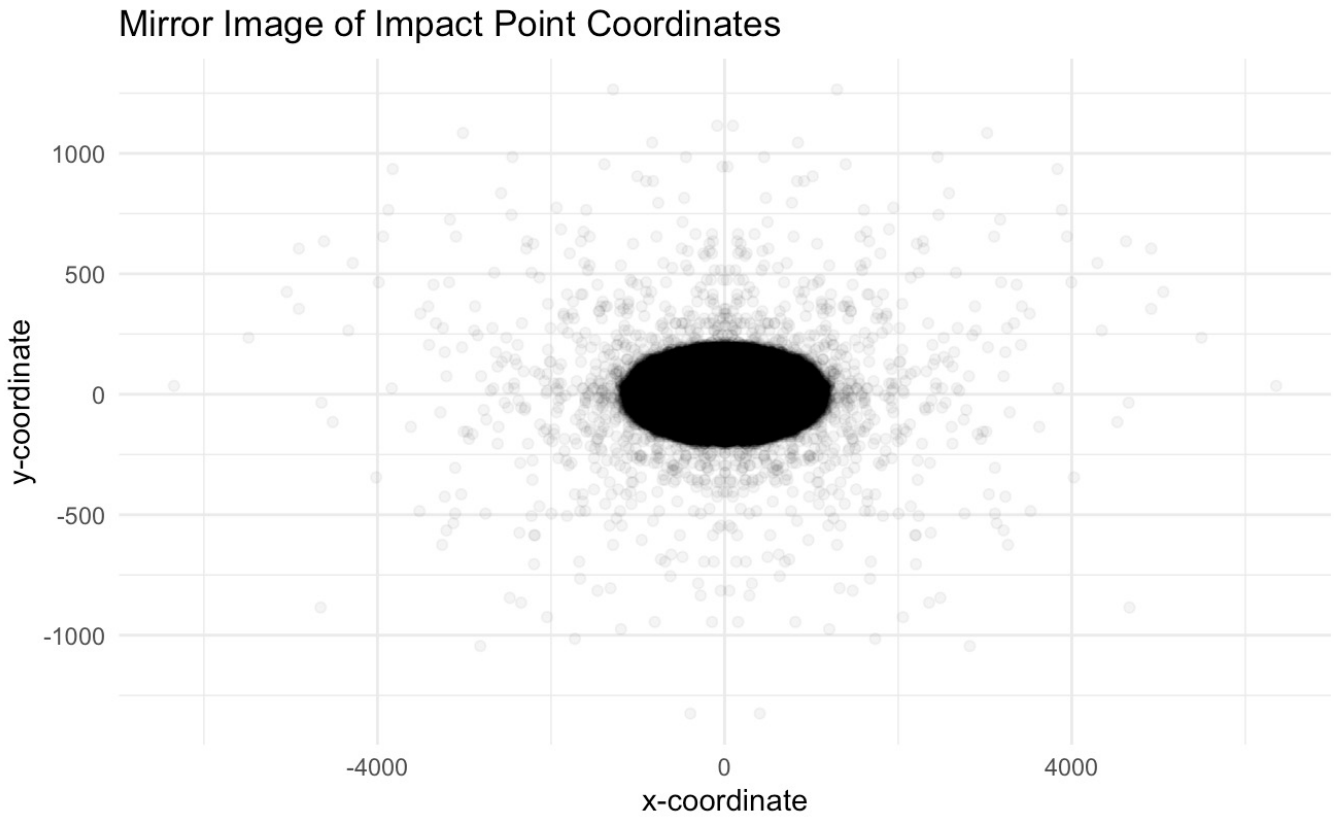
When duplicated the impact points, the distribution resembles gaussian

→ Fit a Gaussian model to data

3. Methodology

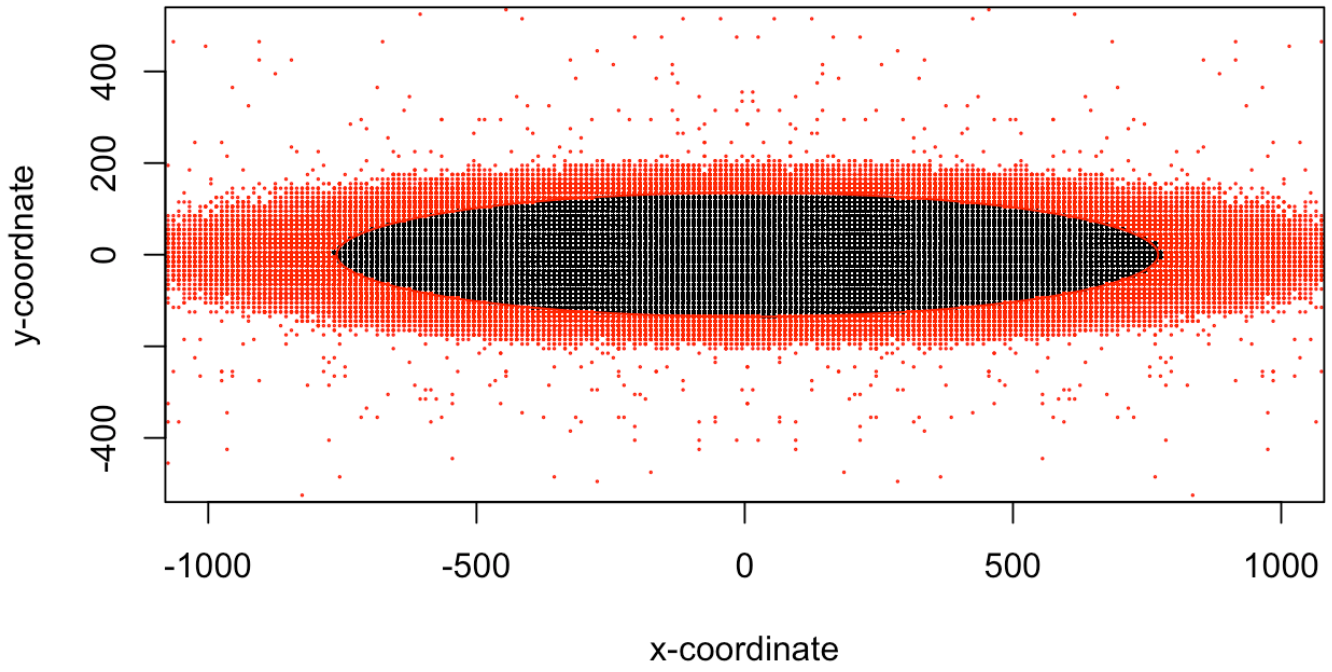
Step 1 :	Create a Gaussian Mixture Model by analyzing the data and its mirrored version
Step 2 :	Find an ellipse that leaves out 100*θ% of data points using the model's parameters
Step 3 :	Bootstrap B ellipses, each excluding 100*θ% of data points Estimate the exclusion boundary with γ% confidence and θ probability

Step 1. Create a Gaussian Mixture Model by analyzing the data and its mirrored version



Density estimates are produced for each data point using a Gaussian mixture model

Step 2. Find an ellipse that leaves out 100*θ% of data points using the model's parameters



- **Black points :** non-excluded data points
- **Red points :** excluded data points that comprise 5% of the total data points
- **Red ellipse boundary :** ellipse that excludes 5% of the data points

Step 3. Estimate the exclusion boundary with γ% confidence and θ probability

A. Resample the data using bootstrap (with replacement) based on the GMM model.

B. Extract mean and covariance matrix parameters for each resampled data set.

C. Calculate the trace of each covariance matrix in the resampled data.

D. Determine the ellipse size using the γ% quantile of the eigenvalues.

E. Calculate the ellipse radius based on the exclusion probability θ.

F. Create ellipses representing the exclusion boundary with a confidence interval of γ and the exclusion probability θ.

4. Results

Simulation Steps:

A. Specify exclusion probability θ and confidence level γ%

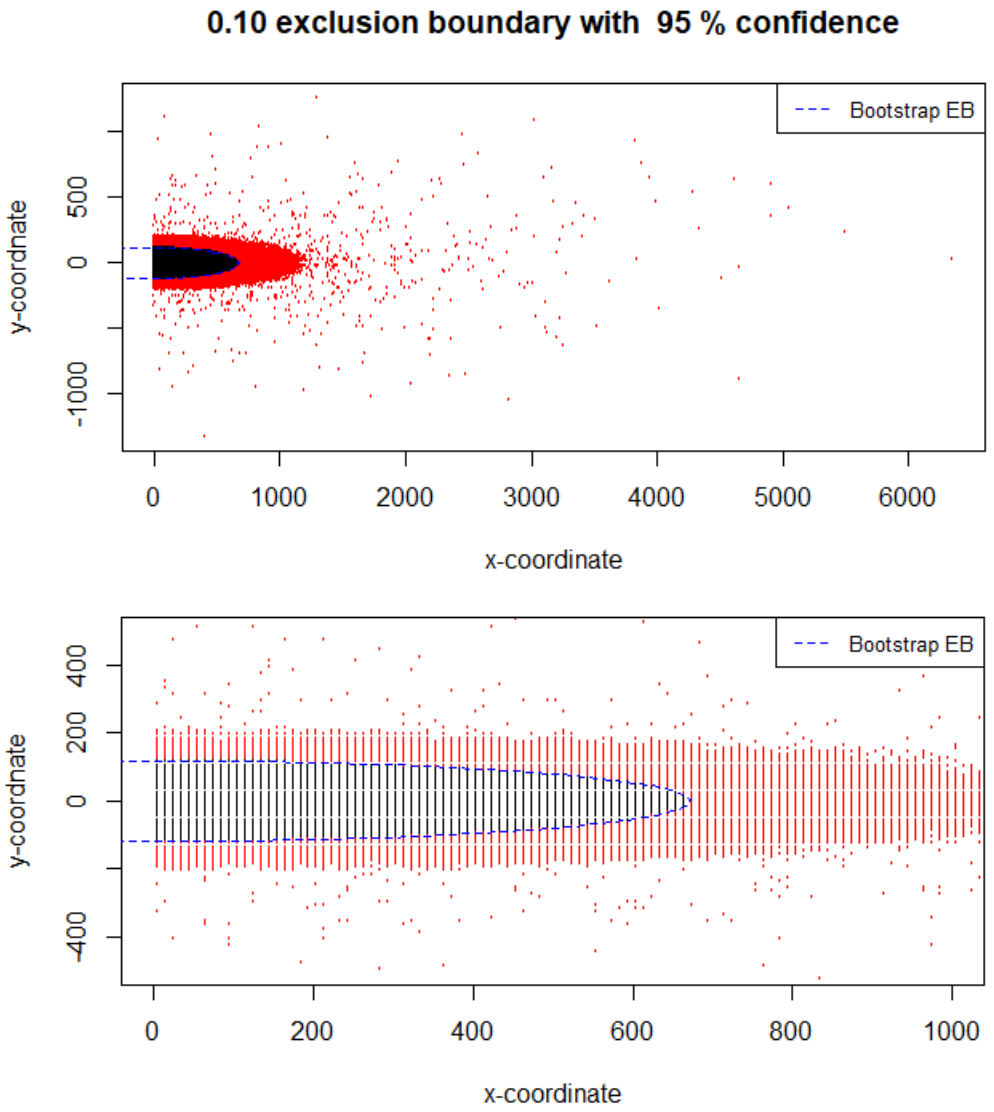
B. Using the methodology from the previous section, construct an exclusion boundary that has probability θ of crossing with γ% confidence

C. Take 100 bootstrap samples of size N from the data, and compute the percentage of points in each sample that lies outside of the exclusion boundary

D. Count the number of bootstrap samples which excludes at most 100*θ*N data points to the pre-determined confidence level γ

E. If our method is exact, there should be (1- γ%)*100 bootstrap samples that have more than 100*θ*N data points that lie outside of the exclusion boundary

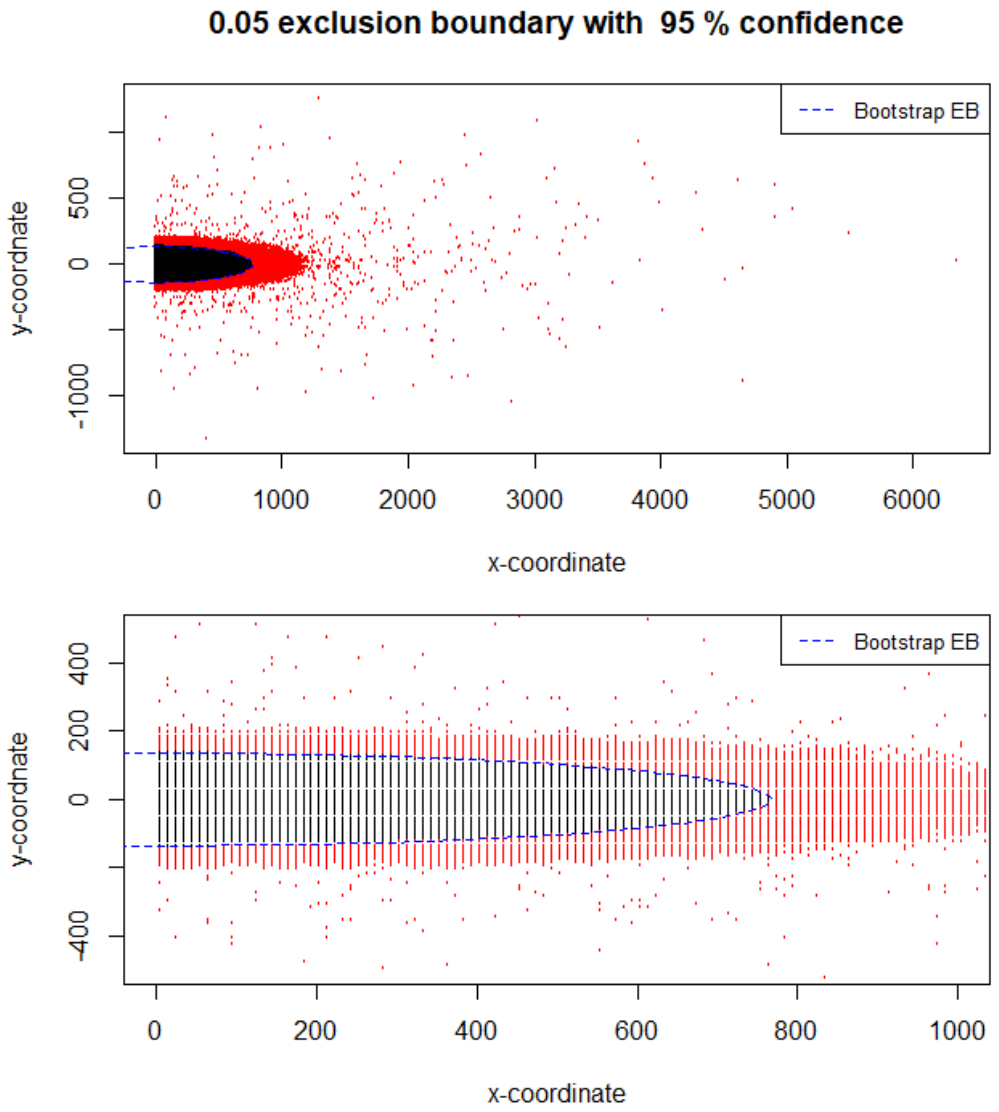
4. Results



N	Number of Unsatisfactory Bootstrap Samples	Mean Exclusion (10%)	Standard Deviation of Exclusion
10,000	27 / 100	9.8035%	0.2998%
100,000	3 / 100	9.8283%	0.0842%
1,000,000	0 / 100	9.8232%	0.0317%

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

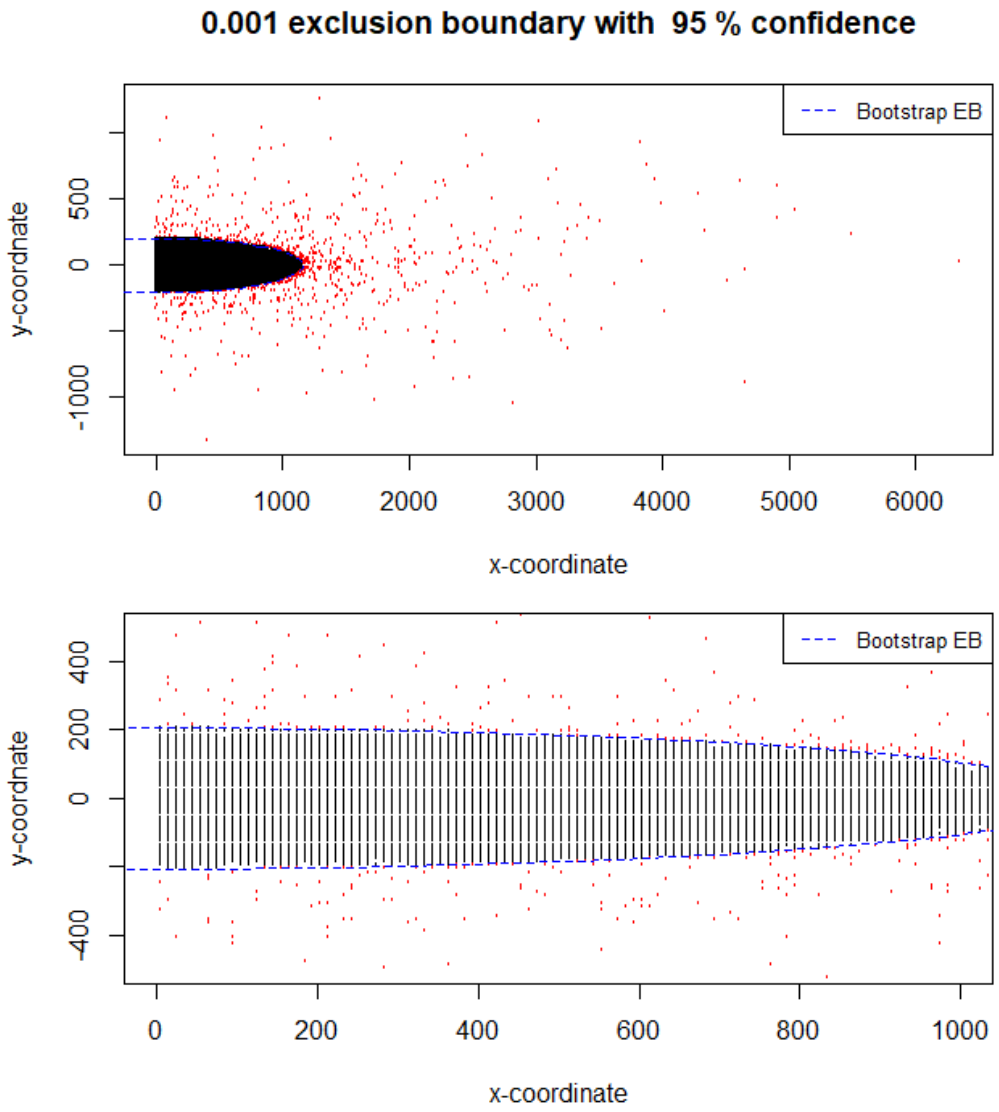
4. Results



N	Number of Unsatisfactory Bootstrap Samples	Mean Exclusion (5%)	Standard Deviation of Exclusion
10,000	34 / 100	4.9098%	0.2167%
100,000	8 / 100	4.9199%	0.0620%
1,000,000	0 / 100	4.9278%	0.0200%

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

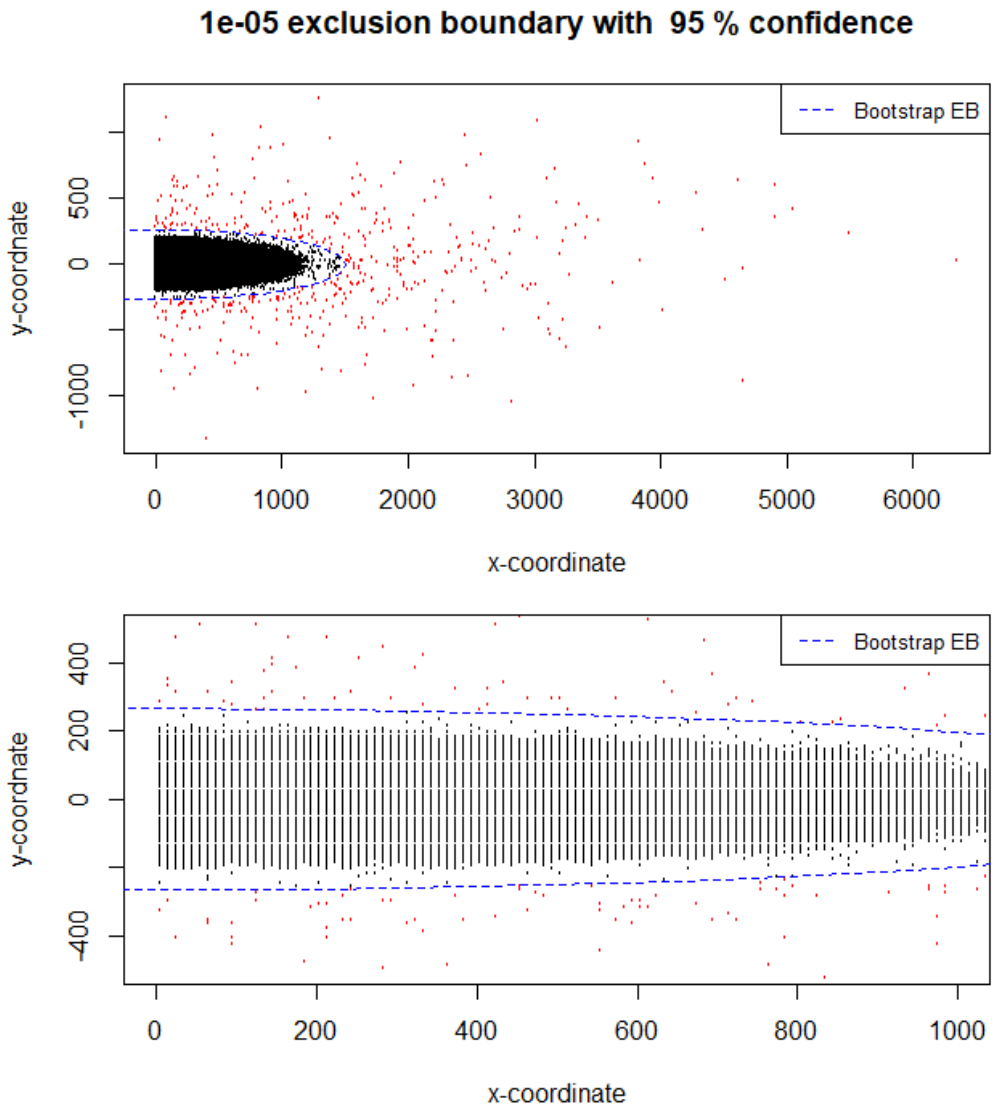
4. Results



N	Number of Unsatisfactory Bootstrap Samples	Mean Exclusion (0.1%)	Standard Deviation of Exclusion
10,000	26 / 100	0.0856%	0.0297%
100,000	7 / 100	0.0876%	0.0001%
1,000,000	0 / 100	0.0884%	0.0000%

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

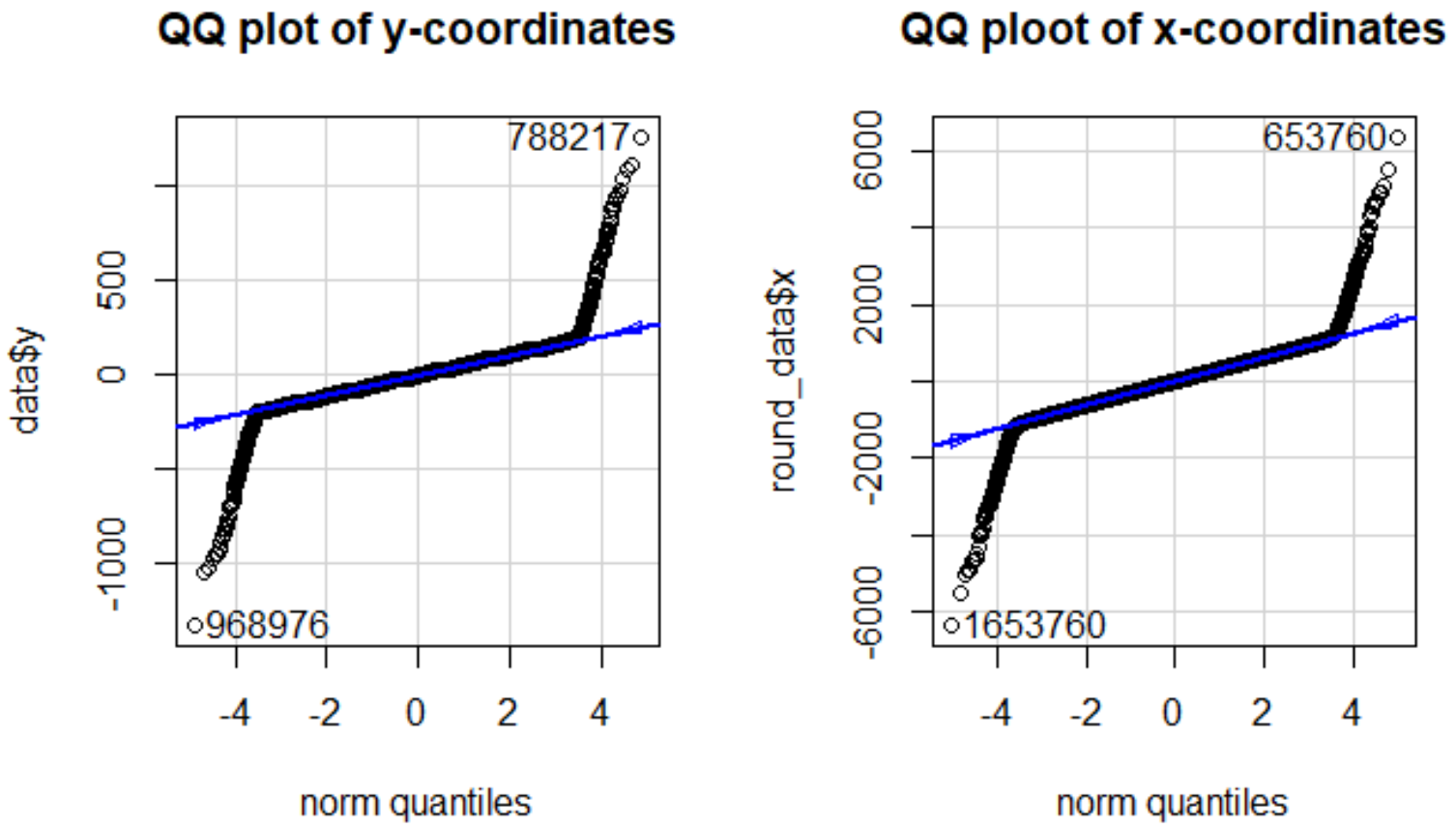
4. Results - Concern



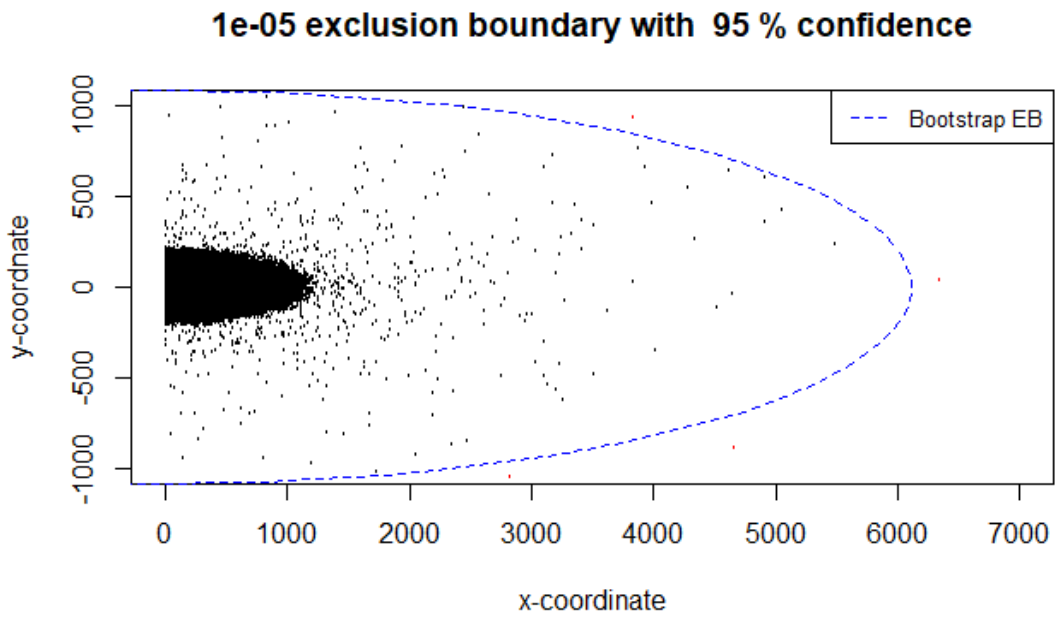
N	Number of Unsatisfactory Bootstrap Samples	Mean Exclusion (0.001%)	Standard Deviation of Exclusion
10,000	100 / 100	0.0453%	0.0172%
100,000	100 / 100	0.0462%	0.0001%
1,000,000	100 / 100	0.0463%	0.0000%

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

4. Results – Adjustment for Extremely Small Theta



4. Results - Concern



N	Number of Unsatisfactory Bootstrap Samples	Mean Exclusion % (0.001% Expected)	Standard Deviation of Exclusion %
10,000	8 / 100	0.0009%	0.0032%
100,000	21 / 100	0.0009%	0.0010%
1,000,000	23 / 100	0.0008%	0.0003%

5. Conclusion

- Our method of identifying an exclusion boundary utilizes the distribution of the data by transforming it to a more familiar distribution and fitting the well-known Gaussian Mixture Model.
- The method can provide an exclusion boundary that has θ probability of being crossed with γ% confidence.
- In the case of extremely small θ, the method performs poorly due to the presence of many extreme observations in the data and relatively small data size. We recommend collecting a sufficiently large sample when θ of interest is extremely small in order to identify an exclusion boundary with high levels of confidence.

5. Conclusion

- Our method uses familiar statistical techniques to create an exclusion boundary.
- The boundary has a specific probability of being crossed and a confidence level.
- When dealing with extremely small probabilities, the method may not perform well due to extreme data points and a small dataset.
- To ensure accurate results with high confidence, gather a larger sample when working with very small probabilities.

THANK YOU!

Alaattin Isilak, Kyu Min Shim, Pamalka Bandara, Sohoon Youn

Team Exclusion Boundary Hunters