# Hawkes process modeling of COVID-19 with mobility leading indicators and spatial covariates

## Overall Framework

## Dependencies

Matlab 2018b

## Quick run

### Use MLE estimate shape and scale parameters for Weibull distribution

**HawkPR('./input_data/NYT_Dconfirmed.csv', './input_data/GoogleMobi_Dconfirmed.csv', './input_data/Demo_Dconfirmed.csv', 14, '', '', 200, 7, 100, './output/mdl.mat', './output/pred.csv')**

### Specify shape and scale parameters for Weibull distribution

**HawkPR('./input_data/NYT_Dconfirmed.csv', './input_data/GoogleMobi_Dconfirmed.csv', './input_data/Demo_Dconfirmed.csv', 14, 8, 4, 200, 7, 100, './output/mdl.mat', './output/pred.csv')**

## Function parameters

### HawkPR( InputPath_report, InputPath_mobility, InputPath_demography, Delta, Alpha, Beta, EMitr, DaysPred, SimTimes, OutputPath_mdl, OutputPath_pred)

| Functiona parameter | Description |
|---|---|
| InputPath_report | Input path for COVID daily report |
| InputPath_mobility | Input path for mobiblity report. |
| InputPath_demography | Input path for spatial demographic features. |
| Delta | Days lagged for mobility. |
| Alpha | Shape parameters for Weibull distribution. Leave it blank string as '' to allow MSE estimation. |
| Beta | Scale parameters for Weibull distribution. Leave it blank string as '' to allow MSE estimation. |
| EMitr | Maximum iterations for EM algorithm |
| DaysPred | Number of days to make prediction. |
| SimTimes | Simulation times for Hawkes processes. Nota that the prediction is the average of number of simulated events among all simulations. |
| OutputPath_mdl | Output path for the trained model. |
| OutputPath_pred | Output path for prediction results. |

## Input Data format

### COVID daily report

- In csv file format. The header should contain "FIPS,State,County,x2020-02-15, ..."
- Date format is in x + 4 digits year + 2 digits month + 2 digits day, i,g., x2020-02-15.
- Each row is a covid daily report for each county.
- The total number of rows is the number of counties.

### Mobility indices

- In csv file format. The header should contain "FIPS,State,County,Type,x2020-02-15 ..."
- Date format is in x + 4 digits year + 2 digits month + 2 digits day, i,g., x2020-02-15.
- Each row is a mobility indices for each county for each type of mobility.
- The total number of rows is the number of counties X number of mobility types

### Spatial demographic features

- In csv file format. The header should contain "FIPS,State,County,Feature 1,Feature 2, ..."
- Each row is demographic features for each county.
- The total number of rows is the number of counties.

## Examples of training data in './input_data/'

| Functiona parameter | Description |
|---|---|
| NYT_Dconfirmed.csv | COVID daily confirmed cases collected by The New York Times (counties with confirmed cases >= 10) |
| NYT_Ddeath.csv | COVID daily deaths collected by The New York Times (counties with deaths >= 1) |
| GoogleMobi_Dconfirmed.csv | Google mobility of counties in NYT_Dconfirmed.csv (Imputated) |
| GoogleMobi_Ddeath.csv | Google mobility of counties in NYT_Ddeath.csv (Imputated) |
| Demo_Dconfirmed.csv | Demographic features of counties in NYT_Dconfirmed.csv |
| Demo_Ddeath.csv | Demographic features of counties in NYT_Ddeath.csv |

## Examples of output data in './output/'

### mdl.mat

- model is save in mat files. (load('mdl.mat') can load the models in MATLAB)
- variable 'mdl' is the summary of Poisson regression model including coefficients
- variable 'mus' is the background rate (Exogenous infections)
- variable 'K0' is the estimated reproduction number
- variable 'alpha, beta' are the esimtated shape and scale parameter for the Weibull distribution

### pred.csv

- In csv file format. The header should contain "FIPS,State,County,x2020_05_21 ..."
- Date format is in x + 4 digits year + 2 digits month + 2 digits day, i,g., x2020_05_21.
- Each row is a mobility indices for each county for each type of mobility.
- There are 'DaysPred' days of prediction starting from the day after the last day in COVID report.