Counterfactual Temporal Point Processes

Pre-requisites

This code depends on the following packages:

1. `networkx`
2. `numpy`
3. `pandas`
4. `matplotlib`

   to generate map plots:

5. `GeoPandas`
6. `geoplot`

In order to install the project dependencies you can run the following command:

```
pip install -r requirements.txt
```

Code structure

- **src/counterfactual_tpp.py:** Contains the code to sample rejected events using the superposition property and the algorithm to calculate the counterfactuals.
- **src/gumbel.py:** Contains the utility functions for the Gumbel-Max SCM.
- **src/sampling_utils.py:** Contains the code for the Lewis' thinning algorithm (`thinning_T` function) and some other sampling utilities.
- **src/hawkes/hawkes.py:** Contains the code for sampling from the hawkes process using the superposition property of tpps. It also includes the algorithm for sampling a counterfactual sequence of events given a sequence of observed events for a Hawkes process.
- **src/hawkes/hawkes_example.ipynb:** Contains an example of running algorithm 3 (in the paper) for both cases where we have (1) both observed and un-observed events, and (2) the case that we have only the observed events.
- **ebola/graph_generation.py:** Contains code to build the Ebola network based on the network of connected districts.
- **ebola/dynamics.py:** Contains code for sampling counterfactual sequence of infections given a sequence of observed infections from the SIR porcess (the `calculate_counterfactual` function). The rest of the code simulates continuous-time SIR epidemics with exponentially distributed inter-event times.

The directory **ebola/data/ebola** contains the information about the Ebola network adjanceny matrix and the cleaned ebola outbreak data.

The directory **ebola/map/geojson** contains the geographical information of the districts studied in the Ebola outbreak dataset. The geojson files are obtained from Nominatim.

The directory **ebola/map/overall_data** contains data for generating the geographical maps in the paper, and includs the overall number of infection under applying different interventions.

The directories **src/data_hawkes** and **src/data_inhomogeneous** contain observational data used to generate Synthetic plots in the paper. You can use this data to re-generate paper's plots. Otherwise, you can simply generate new random samples by the code (which is commented in the corresponding notebooks).

Experiments

Synthetic

- **Inhomogeneous Poisson Processes:** `src/inhomogeneous_experiments.ipynb`
- **Hawkes Processes:** `src/hawkes_experiments.ipynb`

Epidemiological

- **Ebola Epidemic Simulation and Counterfactual Calculations:** `ebola/ebola_experiments.ipynb`
- **Generate Geographical Distribution of infections:** `ebola/map/generate_geopands_data.ipynb`

Execution run-time

In what follows, we report detailed run time of the submitted code on a machine equipped with 48 Intel(R) Xeon(R) 3.00GHz CPU cores and 1.5 TB memory.

Regarding ebola/ebola experiments.ipynb, it takes ~1 minute to generate Figure 1(a), ~6 minutes to generate each of the plots in Figure 1(b) and 55 minutes to generate Figure 3 (note that it combines 40 different experiments.). Regarding src/hawkes experiments.ipynb, it takes ~3 minutes to generate each of the plots in Figures 6 and 7 using the pre-saved data, which is available in supplementary materials. Regarding src/inhomogeneous experiments.ipynb, it takes ~30 minutes to generate each row of plots in Figure 5.