ICT for Smart Mobility

Exercise – Transport modelling
Creating trip generation and trip distribution model of demand

In this practical activity, you will create a demand model for **trip generation** and **trip distribution** (OD matrices),

- The available data consists in a sample of **collected trips by shared cars** in the time window 11:00 pm - 11:59 pm in the city of Amsterdam.
- We are interested in creating a **demand model** for these trips in that time window. You will need to choose an appropriate **zoning system** for the city and produce a **trip generation** and **trip distribution** model for the current scenario. The model you create will be then integrated by someone else with demand models of other means and then used to compare the current scenario with possible routes of new night buses.
- The demand model should generalize and you will perform the **calibration** and **validation** process.

Notes
- We are not interested in the total volume, only in the trip generation and trip distribution.
- We are not interested in segmentation or split by means.
- We are not interested in route assignment.

Data
The dataset “Amsterdam_dataset.csv” is in the folder “Exercises/Transport Modelling/”

Each line corresponds to a collected trip.
The useful columns are:
- ‘start_longitude’: origin O longitude (in EPSG:4326)
- ‘start_latitude’: origin O latitude (in EPSG:4326)
- ‘end_longitude’: destination D longitude (in EPSG:4326)
- ‘end_latitude’’: destination D latitude (in EPSG:4326)
- ‘start_longitude_utm’: origin O longitude (in UTM and then converted to m)
- ‘start_latitude_utm’: origin O latitude (in UTM and then converted to m)
- ‘end_longitude_utm’: destination D longitude (in UTM and then converted to m)
- ‘end_latitude_utm’’: destination D latitude (in UTM and then converted to m)

STEP A - Zoning

Explore the sample locations in the dataset:
- Plot starting and ending points on a map

Propose a zoning system of zones z in Z:
- Use simple uniform regular tilings with square zones with sides of x degrees/meters
- You might only consider zones where you have at least one starting or ending sample (O or D)

STEP B – Trip generation

Given the zoning system Z, propose a trip generation model:

- Estimate the probability of origins P(O) and destinations P(D).
- The calibrated model from the samples is the one that assigns to P(O) the fraction of observed trips starting from O.
    - Evaluate the total log likelihood metric to be maximized: 1/N sum_(samples s) log(P(s))
- Show the distributions P(O) and P(D) of the trip generation model on a map.

STEP C – Trip generation validation

Now observe the performance of the created model on ‘new’ data. Indeed, the model might overfit the observed sample.

- Divide the sample dataset into random training (2/3) and validation (1/3) samples.
- Now go back to STEP B and recompute the trip generation on the training.
- Evaluate the total log likelihood of the training and validation. Which one is worse? Why?
- Now go back to STEP A: what happens if you change the zone size? If you choose many zones, can you trust these data?
- Plot the total log likelihood for training and validation data with respect to the number of zones (or size of zoning system). What is a good trade-off for zoning in your opinion?

STEP D – Trip distribution

Given the zoning system Z, propose a trip distribution model:

- Estimate the full OD matrix. i.e. the probability of moving from one origin O to a destination D: P(O,D)
- The calibrated model from the samples is the one that assign to P(O,D) the fraction of observed trips starting from O and going to D
    - Evaluate the total log likelihood metric to be maximized: 1/N sum_(samples s) log(P(O(s), D(s)))
- How many intra zones trip do you have?
- Given two different origins, show the distribution of destinations on a map.
- Given two different destinations, show the distribution of origins on a map.

STEP E – Trip distribution validation

Repeat STEP C for trip distribution.