Post lecture questions Demand Modelling

Answers post-lecture questions Trip generation

Trip generation: theory

(a) What are the outputs of trip generation models?

The outputs are the number of departures from a certain zone (trip production) and the number of arrivals at a certain zone (trip attraction).

(b) Name four trip purposes that are typically distinguished in trip generation models.

- work
- education
- leisure
- shopping
- business

(c) Why is the number of trips usually calculated *per trip purpose*?

Each trip purpose may have different characteristics and therefore different models may be needed for each trip purpose. Also, in the trip distribution and mode choice models, a distinction per trip purpose is usually important, therefore the number of trips per purpose has to be known from the trip generation step.

(d) Name three different models for determining the trip generation.

- regression models
- cross-classification models
- discrete choice models

(e) What is meant by trip balancing?

With trip balancing, the trip production and trip attraction are made equal.

In trip generation, regression models, cross-classification, and binary logit models can be used to estimate the number of trips produced by a household.

Assume that there has been a household survey regarding the number of trips made per day. Households were asked questions about their household income and the number of cars available. The following linear regression model was estimated from the survey data:

$T = 1.75 + 0.00002 I + 0.5 A$,

where

T = number of trips made

I = household income

A = number of cars available

The number of trips is linear with the number of available cars in the household, which is not very realistic for households with more than 2 cars. One way of solving this problem is to introduce dummy variables in the regression model.

(a) Write down a regression model with dummy variables such that the number of trips is no longer linearly dependent on the number of cars.

Nonlinearity has to be introduced into the regression model such that it is no longer linear in the number of available cars. For example, the regression model may look as follows:
$T = \beta_0 + \beta_1 I + \beta_2 A_1 + \beta_3 A_2 + \beta_4 A_3,$
where

$$A_1 = \begin{cases} 1, & \text{if 1 car available} \\ 0, & \text{otherwise} \end{cases},$$

$$A_2 = \begin{cases} 1, & \text{if 2 cars available} \\ 0, & \text{otherwise} \end{cases}, \text{ and}$$

$$A_3 = \begin{cases} 1, & \text{if 3 or more cars available} \\ 0, & \text{otherwise} \end{cases}$$

Now there is not just one parameter to estimate for the car attribute, but instead there are three (different) parameters, making it a nonlinear relationship between the number of trips and the number of cars available.

The same survey data can be used in a cross-classification model by classifying households into income groups and car ownership. In the table below this classification is shown. In each cell the average value of the trips made by that household type is computed. The number of observations is stated between brackets.

income / cars available	0 cars	1 car	≥2 cars
< € 15.000	1.9 (7)	2.4 (5)	?? (0)
€ 15.000 - € 30.000	2.2 (12)	2.8 (25)	2.9 (3)
> € 30.000	2.3 (2)	3.1 (11)	3.5 (9)

A disadvantage of the cross-classification model is that a relatively large survey is necessary to have sufficient observations in each cell. As can be seen in the table, there are no observations for households with low income and 2 or more cars. Using multiple class analysis (MCA), this problem can be resolved.

(b) Using MCA, compute the estimate for the average number of trips made by low income households with 2 or more cars.

First, compute the overall average:
(1.9*7+2.4*5+2.2*12+2.8*25+2.9*3+2.3*2+3.1*11+3.5*9) / (7+5+12+25+3+2+11+9) = 2.71
Average income < € 15.000: (1.9*7+2.4*5) / (7+5) = 2.11
Deviation from average: 2.11 – 2.71 = -0.6
Average ≥2 cars: (2.9*3+3.5*9) / (3+9) = 3.35
Deviation from average: 3.35 – 2.71 = 0.64
Estimate for low income households with 2 cars or more:
2.71 – 0.6 + 0.64 = 2.75 trips

In a binary logit model, the trip generation is determined by sequentially computing the probability of making additional trips. Suppose that the binary logit model is applied 4 times successively to compute the probability for an individual of making additional trips, given by the following numbers:

- The probability of a person to make one or more trips is 90%.
- For people who decide to make one or more trips, the probability of making one trip is 50% (such that the probability of making two or more trips is 50%).
- For people who decide to make two or more trips, the probability of making two trips is 80% (such that the probability of making three or more trips is 20%).
- For people who decide to make three or more trips, the probability of making three trips is 100% (such that the probability of making 4 or more trips is 0%).

(c) Determine the average number of trips made by an individual using the outcomes of the binary logit model.

Probability of making no trips: P(0) = 0.1
Probability of making 1 trip:
P(1) = P(1+) * P(1) = 0.9 * 0.5 = 0.45
Probability of making 2 trips:
P(2) = P(2+) * P(2) = (0.9 * 0.5) * 0.8 = 0.36
Probability of making 3 trips:
P(3) = P(3+) * P(3) = (0.9 * 0.5 * 0.2) * 1 = 0.09

Average number of trips being made =
0 * P(0) + 1 * P(1) + 2 * P(2) + 3 * P(3) = 1.44 trips

Answers post-lecture questions Trip distribution

Formulate the direct demand model for trip distribution and define its variables.

$$T_{ij} = \rho Q_i X_j F_{ij}$$

T_{ij} = number of trips from i to j

ρ = measure of average trip intensity

Q_i = production potential of zone i

X_j = attraction potential of zone j

F_{ij} = accessibility of zone j from i

Suppose a city would like to build more houses in a certain suburb. This implies that more road infrastructure is needed for all residents in that suburb.
Which trip distribution model can be best applied for computing the OD trip matrix?

A model based on the gravity model (either origin-based singly constrained or doubly constrained).
Not the growth factor model, since it cannot capture new travel patterns for new residential areas (it is based on historical data).

Consider two zones with the following data:

	number of inhabitants	number of jobs
Zone A	1000	300
Zone B	800	200

The number of inhabitants has been determined with a higher precision than the number of jobs. On average, the number of departing trips is 0.25 per inhabitant, and the number of arriving trips is 0.8 per job.

All the travel resistances (intra-zonal and inter-zonal) may be assumed equal.

Determine the trip distribution.

...	...	250
...	...	200
240	160	

1	1	250
1	1	200
270	180	

125	125	250
100	100	200
270	180	

150	100	250
120	80	200
270	180	

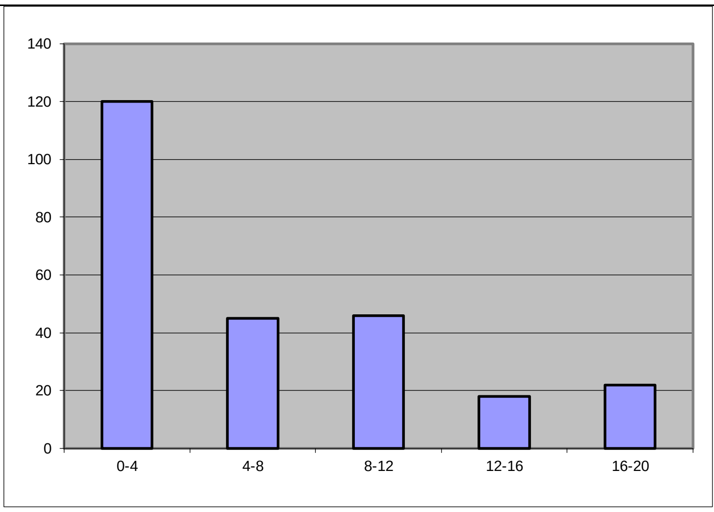
Consider a transport network with four zones. The distances (in km) between the zones are given in the skim matrix below. Also a measured origin-destination matrix is provided.

skim	zone 1	zone 2	zone 3	zone 4
zone 1	1	2	5	19
zone 2	3	1	10	15
zone 3	7	9	1	9
zone 4	18	17	6	1

trips	zone 1	zone 2	zone 3	zone 4
zone 1	20	16	12	7
zone 2	15	31	26	18
zone 3	18	8	22	12
zone 4	11	4	15	16

(a) Compute and sketch the trip length frequency distribution using trip length intervals of 4km.

0 – 4 km: 20 + 31 + 22 + 16 + 16 + 15 = 120
4 – 8 km: 12 + 15 + 18 = 45
8 – 12 km: 8 + 22 + 12 = 46
12 – 16 km: 18
16 – 20 km: 11 + 4 + 7 = 22



(b) What is the difference between a trip length frequency distribution and a trip distribution function?

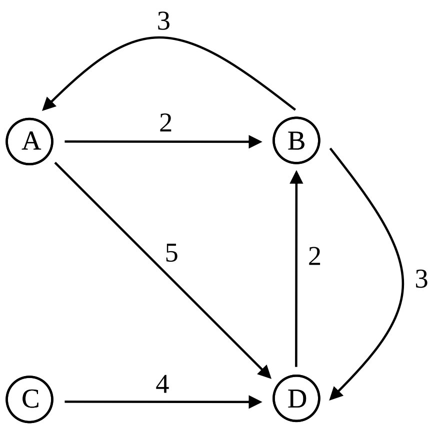
A trip distribution describes *behaviour*; it states what is the accessibility and therefore the willingness to make a trip for each distance. A trip length frequency distribution describes the absolute amount of trips being made for each distance class, and it network specific.

(c) The trip distribution function is unknown. Explain in words how you can estimate the trip distribution function using the given data.

We know:
- the productions and attractions, from the observed trip matrix
- from the skim matrix, the observed trip length frequency distribution

Using the Poisson estimator, we can estimate the trip distribution function. Start with a matrix with all zeros. Scale the matrix towards the productions and attractions. Then, update the distribution function values using the know trip length frequency distribution. Again, perform the scaling, etc.

Consider the following transportation network between four cities (A, B, C, and D), with the distances (in km) denoted on the (directed) links. The trip production for cities A and D is 100 trips, and for B and C 200 trips. The trip attraction for cities A and D is 150 trips, and for city B 300 trips. No trips are attracted to zone C. No intrazonal trips are considered.



(a) Determine the skim matrix with travel distances between all cities.

$$c = \begin{bmatrix} \infty & 2 & \infty & 5 \\ 3 & \infty & \infty & 3 \\ 9 & 6 & \infty & 4 \\ 5 & 2 & \infty & \infty \end{bmatrix}$$

Suppose that the trip distribution function is given by $f(c_{ij}) = 1/c_{ij}$, where c_{ij} is the travel distance from city *i* to city *j*.

(b) Using the doubly constrained gravity model, determine the origin-destination trip matrix (perform only one iteration).

The matrix with accessibilities F_{ij} is

	A	**B**	**C**	**D**	**Prod.**
A	0	0.500	0	0.200	100
B	0.333	0	0	0.333	200
C	0.111	0.167	0	0.250	200
D	0.200	0.500	0	0	100
Attr.	150	300	0	150	600

Scaling towards the productions yields

	A	**B**	**C**	**D**	**Prod.**
A	0	71.429	0	28.571	100
B	100	0	0	100	200
C	42.105	63.158	0	94.737	200
D	28.571	71.429	0	0	100
Attr.	150	300	0	150	600

Finally, scaling towards the attractions yields

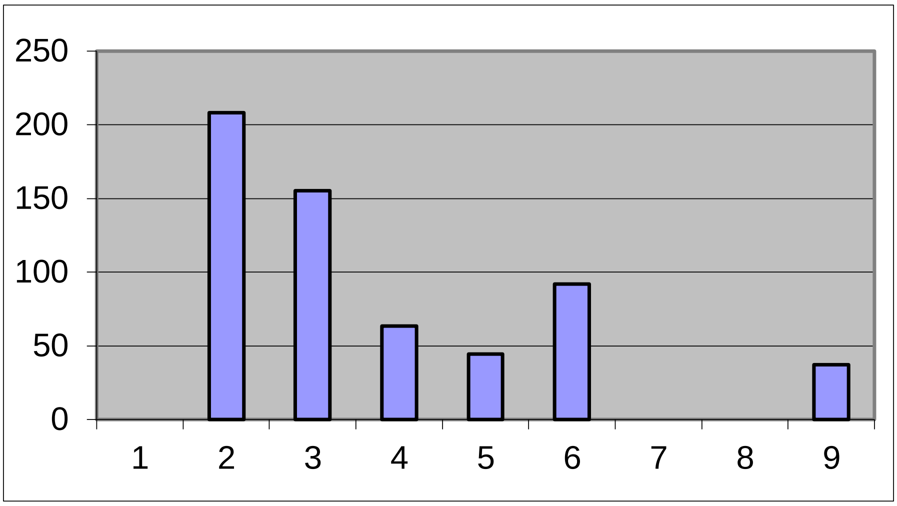
	A	**B**	**C**	**D**	**Prod.**
A	0	104.01	0	19.192	100
B	87.885	0	0	67.172	200
C	37.004	91.971	0	63.636	200
D	25.110	104.015	0	0	100
Attr.	150	300	0	150	600

The above matrix provides the trip matrix after one iteration.
(Note that it is also possible to first scale the columns and then the rows, leading to a slightly different answer)

(c) Sketch the trip length distribution, based on the OD matrix obtained in (b).

Number of trips in different distance classes:
1 km: 0
2 km: 104+104 = 208
3 km: 88+67 = 155

4 km: 64
5 km: 25+19 = 44
6 km: 92
7 km: 0
8 km: 0
9 km: 37



(d) Explain why the shape of the trip length distribution is different from the shape of the trip distribution function.

The trip distribution function describes the behaviour of travellers, assuming all distance classes are available. The trip length distribution considers the possibilities in the network. As such, it describes the travel behaviour constrained to the available distances to destinations in the network.

Suppose that the trip distribution function is unknown, and that only the mean trip length is given.

(e) Which method could be used to perform the trip distribution and estimate the trip distribution function? Explain how this method works.

If only the mean trip length is known, one can use Hyman's method. In this method, an exponential trip distribution function is assumed with a parameter alpha. As a first guess, alpha = 1/MTL, where MTL is the given mean trip length. A trip distribution according to the doubly constrained gravity model is conducted. Then the MTL from the trip matrix resulting from the gravity model is computed and compared to the given MTL. In each iteration, parameter alpha is changed accordingly to approximate the given MTL, until the given MTL is approximated sufficiently.

Answers post-lecture questions Simultaneous distribution and modal split

We can use the gravity model for simultaneously computing the trip distribution and model split. Suppose that the trip productions are known and that the attractions are unknown.

(a) Formulate the (simultaneous) singly-constrained gravity model for computing the OD matrix per mode, and define all its variables and parameters.

$T_{ijm} = a_i P_i X_j F_{ijm}$, where $F_{ijm} = f_m(c_{ijm})$.
T_{ijm} = number of trips from i to j for mode m
P_i = trip production from zone i
X_j = trip attraction potential of zone j
F_{ijm} = accessibility of zone i from zone j for mode m
c_{ijm} = travel cost/impedance from zone i to j for mode m
a_i = scaling/balancing factor
$f_m()$ = trip distribution function for mode m

(b) Name two advantages of a simultaneous trip distribution/modal split model over a sequential model.

Advantages:
- (from a behavioural point of view) It resembles more closely the true choice behaviour, people make destination and mode choices usually simultaneously and not sequentially.
- (from a methodological point of view) There is no need to determine travel costs for all modes together.

In the sequential trip distribution / modal split computation, you have to specify the travel resistance per OD pair in order to compute the distribution. If for a certain OD pair more mode alternatives are available, characterized by their own travel resistances c_{ijv}, how would you express the general (combined) travel resistance c_{ij} for this OD pair? Explain your answer.

If there is a large overlap between the alternatives:

$$c_{ij} = \min_v \{c_{ijv}\}$$

If there is hardly any overlap:

$$c_{ij} = -\frac{1}{\alpha} \ln \sum_v \exp(-\alpha c_{ijv})$$

Consider the following table with distribution values for computing the trip distribution:

	A		B		C		
	car	PT	car	PT	car	PT	
A	67	19	19	0.7	26	7	100
B	19	0.7	67	19	7	19	600
C	26	7	7	19	67	19	400
	200		1200		600		2000

The mode-specific values of the trip distribution function have been generated from:
$F_{ijv} = c \cdot \exp(-t_{ijv})$

(a) What is the PT share on the relation A-C when computing the doubly constrained simultaneous modal split / distribution?

(b) What is the PT share for A-C when performing a sequential trip distribution and modal split, using a logit model for the modal split based on travel times with scale parameter 2?

(a) The PT share remains 7/(7+26) = 21%, even after scaling.

(b) The modal split share can be computed by

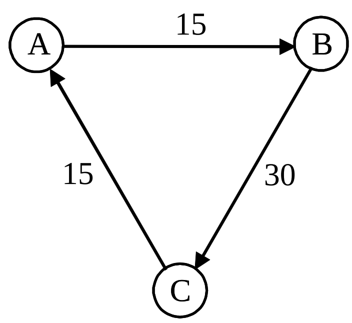
$$p_{\text{pt}} = \frac{\exp(-2t_{\text{pt}})}{\exp(-2t_{\text{pt}}) + \exp(-2t_{\text{car}})}$$

We know

$$\begin{cases} F_{\text{pt}} = c \cdot \exp(-t_{\text{pt}}) = 7 \\ F_{\text{car}} = c \cdot \exp(-t_{\text{car}}) = 26 \end{cases} \Rightarrow \begin{cases} \exp(-2t_{\text{pt}}) = (7/c)^2 \\ \exp(-2t_{\text{car}}) = (26/c)^2 \end{cases}$$

$$p_{\text{pt}} = \frac{(7/c)^2}{(7/c)^2 + (26/c)^2} = 6.7\%$$

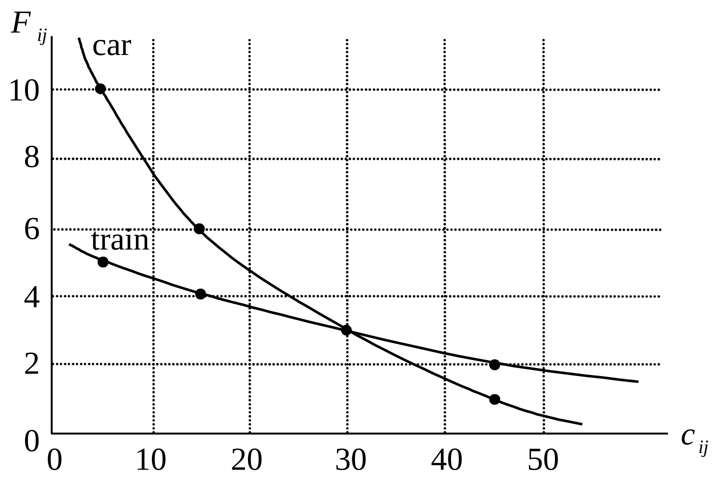
Three zones, A, B, and C, are connected by roads and rails as indicated below. Each (uni-directional) link represents both a road segment as well as a rail segment, and the distances are indicated (for simplicity it is assumed that for each link the length of road and rail are equal).



The trip production and attraction for each zone is given in the table below.

Zone	Production	Attraction
A	2800	2450
B	3600	4260
C	1550	1240

Furthermore, the trip distribution functions for each mode (car and train), which described the accessibilities F_{ij} from zone i to zone j for that mode as a function of the travel distance c_{ij} [km], are indicated in the next figure. Assume that all intrazonal distances equal 5.



(a) Determine the skim matrix $c = [c_{ij}]$ consisting of travel distances.

Determine the travel distances from the 1st figure. Note that all intrazonal distances are assumed equal to 5.

$[c_{ij}]$ =

	A	B	C
A	5	15	45
B	45	5	30
C	15	30	5

(b) Compute the simultaneous trip distribution / modal split based on the doubly constrained gravity model for obtaining the trip matrices for the car and train (perform only 1 iteration).

Use the 2nd figure to convert the skim matrix from (a) into the accessibility matrix. Note that we now distinguish car and train, since the accessibility function is mode specific.

$[F_{ijv}]$ =

	A		B		C	
	car	train	car	train	car	train
A	10	5	6	4	1	2
B	1	2	10	5	3	3
C	6	4	3	3	10	5

Then calculate the (mode specific) origin-destination (OD) matrix based on the accessibility matrix, and perform one iteration. (Note that the productions equal the attractions, such that trip balancing is not needed.)

$[T_{ijv}]$ =

	A		B		C			
	car	*train*	*car*	*train*	*car*	*train*	**factor**	**Production**
A	10	5	6	4	1	2	2800/28	**2800**
B	1	2	10	5	3	3	3600/24	**3600**
C	6	4	3	3	10	5	1550/31	**1550**
Attraction	**2450**		**4260**		**1240**			

	A		B		C			
	car	*train*	*car*	*train*	*car*	*train*	**factor**	**Production**
A	1000	500	600	400	100	200	2800/28	**2800**
B	150	300	1500	750	450	450	3600/24	**3600**
C	300	200	150	150	500	250	1550/31	**1550**
Attraction	**2450**		**4260**		**1240**			

	A		B		C		
	car	*train*	*car*	*train*	*car*	*train*	**Production**
A	1000	500	720	480	63.6	127.2	**2800**
B	150	300	1800	900	286.2	286.2	**3600**
C	300	200	180	180	317.9	159.0	**1550**
factor	2450/2450		4260/3550		1240/1950		
Attraction	**2450**		**4260**		**1240**		

(c) Compute and sketch the trip length frequency distribution for the car (choose your own length classes, use at least three classes).

The skim matrix in (a) shows the travel distance per OD pair, while the OD matrix in (b) shows the number of trips performed per OD pair. By combining these two matrices the trip length frequency distribution can be computed. For example, choose classes of size 20. Then the trip length frequency distribution for the mode car will look as follows:

bin	no. trips	OD-pairs
0 - 20	4137.9	AA,AB,BB,CA,CC
20 - 40	466.2	BC,CB
40 - 60	213.6	AC,BA