

CIE4801 Transportation and spatial modelling

Building OD-matrices & OD-matrix estimation

Rob van Nes, Transport & Planning

17/4/13

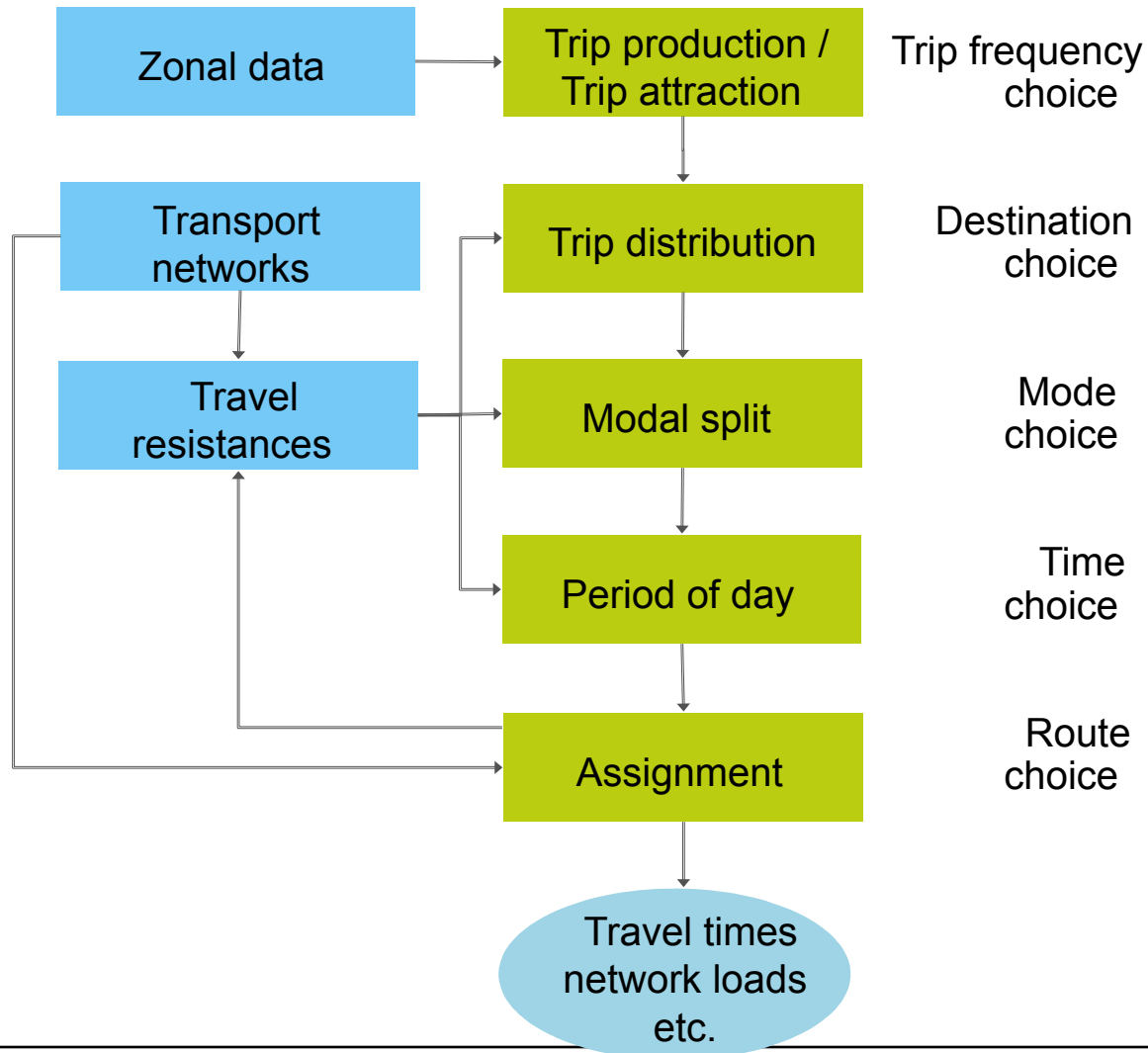
Contents

- Building OD-matrices
 - Choice models
 - Trip generation functions
 - Trip distribution functions
 - (Travel time functions)
- Estimating OD-matrices using counts
- Special topics

1.

Building OD-matrices

Overview framework



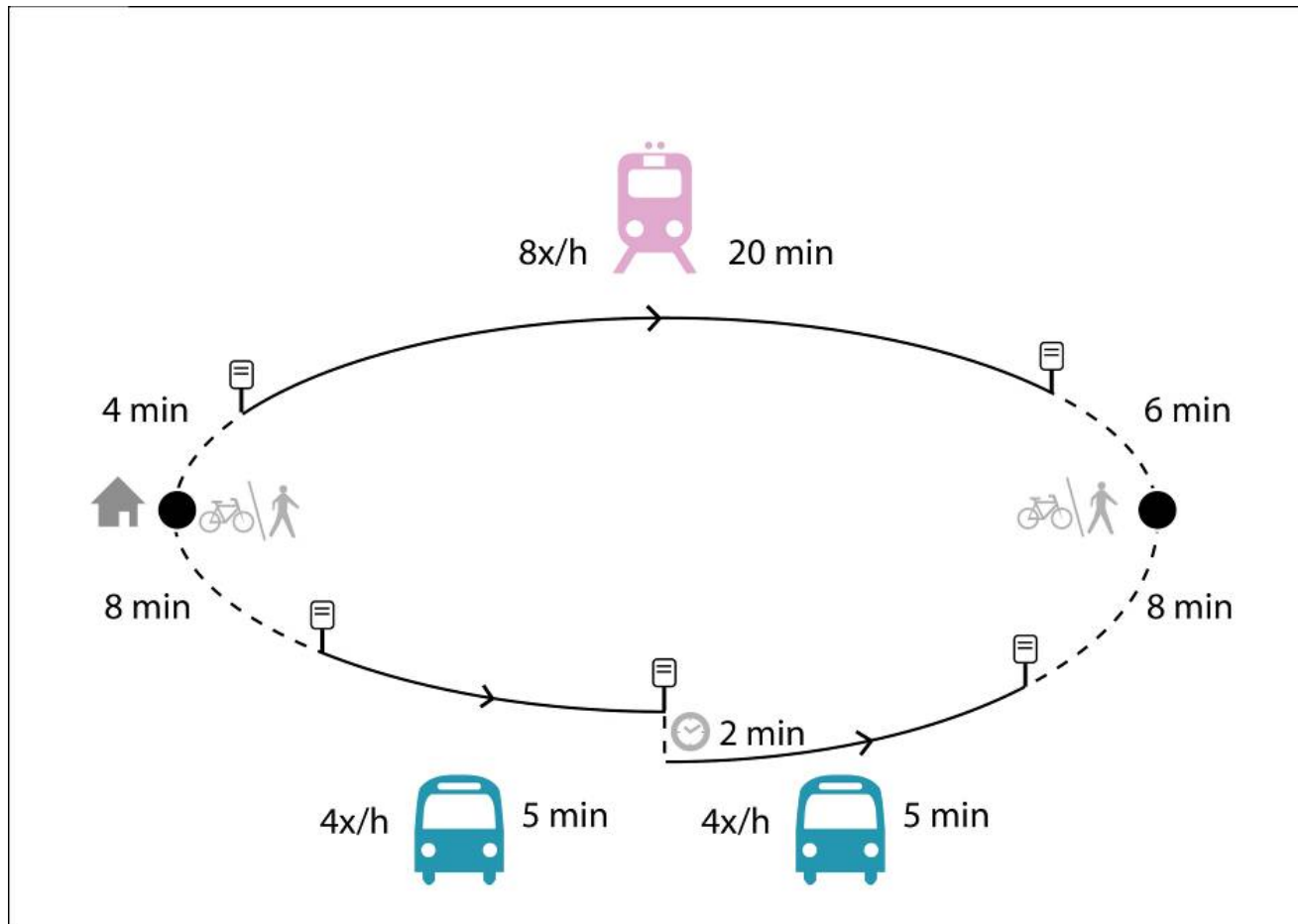
1.1

Building OD-matrices *Choice models*

How to determine choice functions?

- Observe choices as well as the considered alternatives
 - Revealed preference: actual behaviour
 - Stated preference: choice experiments
- Formulate utility functions
- Estimate parameters in utility function such that the probability of the chosen alternative is highest
 - Maximum likelihood method
- Check the quality of the model
 - e.g. ρ^2 ($\approx R^2$) and t-statistics of the parameters
- Dedicated software such as BIOGEME (e.g. see CIE4831)

Example stated choice experiment



Simple example (1 / 3)

Resp	Alternative 1 (chosen)				Alternative 2				Alternative 3			
1	Time	Cost	Delay	Free way	Time	Cost	Delay	Free way	Time	Cost	Delay	Free way
2	30	5	5	1	30	4	6	1	35	3	2	0
3	..											
4												
5												

Simple example (2/3)

Resp	Alternative 1 (chosen)				Alternative 2				Alternative 3			
1	Time	Cost	Delay	Free way	Time	Cost	Delay	Free way	Time	Cost	Delay	Free way
2	30	5	5	1	30	4	6	1	35	3	2	0
3	..											
4												
5												

Specify utility function $V_i = \alpha_t \cdot time_i + \alpha_c \cdot cost_i + \alpha_d \cdot delay_i + \alpha_f \cdot freeway_i$

Assume values for α_x and determine p_{chosen} e.g. $\alpha_x = 1 \Rightarrow p_{chosen}(1) = 0.21$

Simple example (3 / 3)

Likelihood is the product of the probabilities of all chosen alternatives

$$L = \prod_r p_{chosen}(r) \Rightarrow \ln(L) = \sum_r \ln(p_{chosen}(r))$$

r	Alternative 1 (chosen)				Alternative 2				Alternative 3				P ₁	Ln(p ₁)
	Time	Cost	Delay	Free way	Time	Cost	Delay	Free way	Time	Cost	Delay	Free way		
1														
2	30	5	5	1	30	4	6	1	35	3	2	0	0.21	-1.55
3	..													
4														
5														

Use e.g. Solver to determine values for α_x that maximise $\ln(L) = \sum_r \ln(p_1(r))$

1.2

Building OD-matrices
Trip generation

Trip generation functions

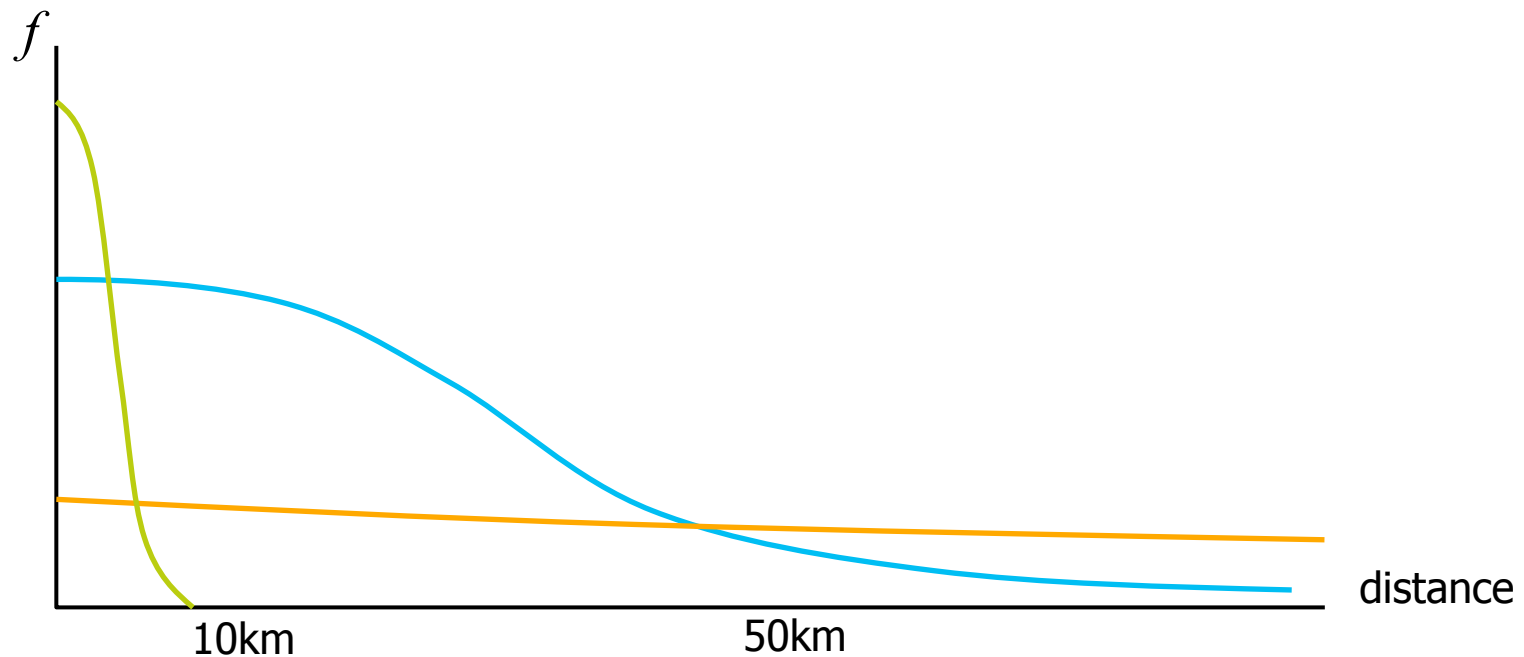
- Linear regression
- Usually at a more aggregate level than zones
 - e.g. Municipality
- Check quality of the model
 - R^2 and t-statistics of the parameters
- See Omnitrans exercise 2

1.3

Building OD-matrices
Trip distribution

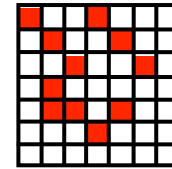
Trip distribution functions

- Trip distribution or deterrence functions can be estimated using the Poisson-estimator



What do you need?

- Observed OD-matrix from a survey
 - Not necessarily complete
 - Usually at an aggregate level (e.g. municipality)
- Cost functions
 - Your definition in time, cost, length plus.....
 - What's the travel time between two cities?
- Discretisation of the cost function $F(c_{ij}) \Rightarrow F_k(c_{ij})$
 - Preferably each class having a similar rate of observations



Mathematical background (1 / 2)

- Key assumption: number of trips per OD-pair is Poisson distributed

- Model formulation: $\hat{T}_{ij} = Q_i \cdot X_j \cdot F_k(c_{ij})$

- Poisson model: $P(x) = \frac{e^{-\lambda} \lambda^x}{x!}$

- For $T_{ij} \Rightarrow p(T_{ij}) = \frac{e^{-(Q_i \cdot X_j \cdot F_k(c_{ij}))} (Q_i \cdot X_j \cdot F_k(c_{ij}))^{T_{ij}}}{T_{ij}!}$

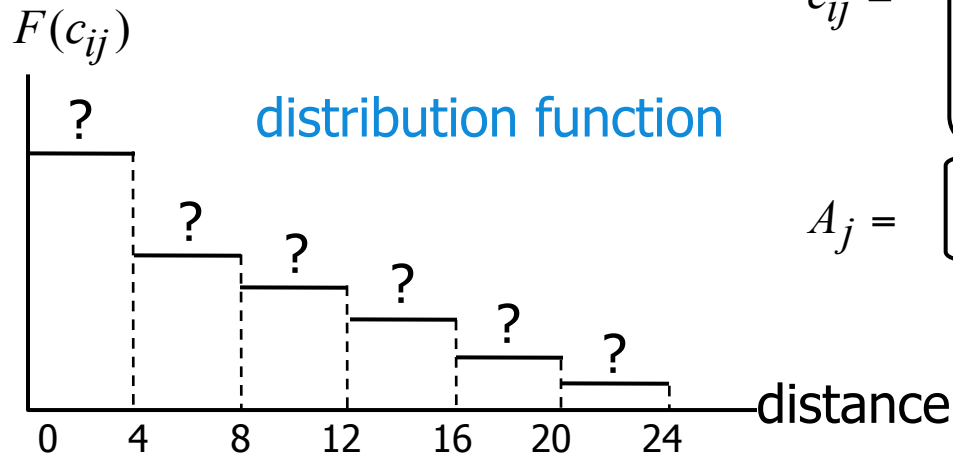
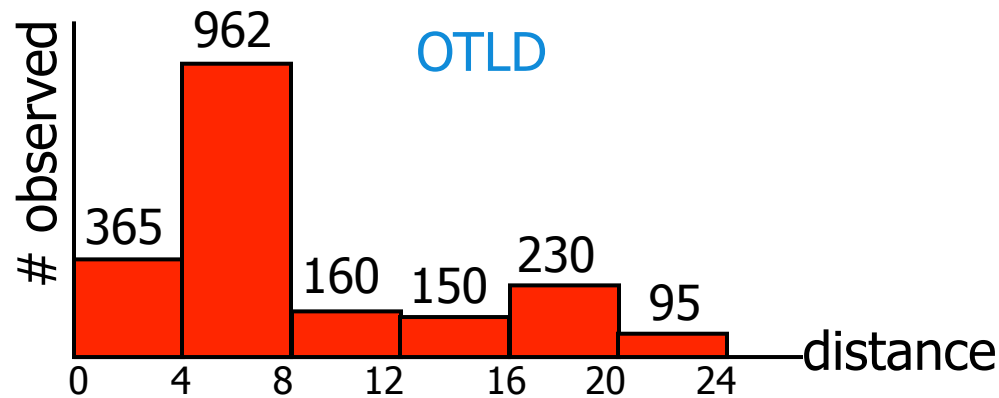
- For a set of N observations n_{ij} the likelihood becomes

$$p(\{n_{ij}\} | Q_i, X_j, F_k(c_{ij})) = \prod_{i,j \in N} \frac{e^{-(Q_i \cdot X_j \cdot F_k(c_{ij}))} (c Q_i \cdot X_j \cdot F_k(c_{ij}))^{n_{ij}}}{n_{ij}!}, c = \frac{\sum_{ij \in N} n_{ij}}{\hat{T}}$$

Mathematical background (2/2)

- Setting derivatives equal to 0 yields 3 linear equations in which each parameter is function of the other two
- Similar solution procedure as for trip distribution:
- Determine the constraints
 - For each origin i the number of observed trips (departures)
 - For each destination j the number of observed trips (arrivals)
 - For each class F_k the number of observed trips
- Set all parameters Q_i , X_j and F_k equal to 1
- Determine successively the values for Q_i , X_j and F_k until convergence

Example Poisson estimator



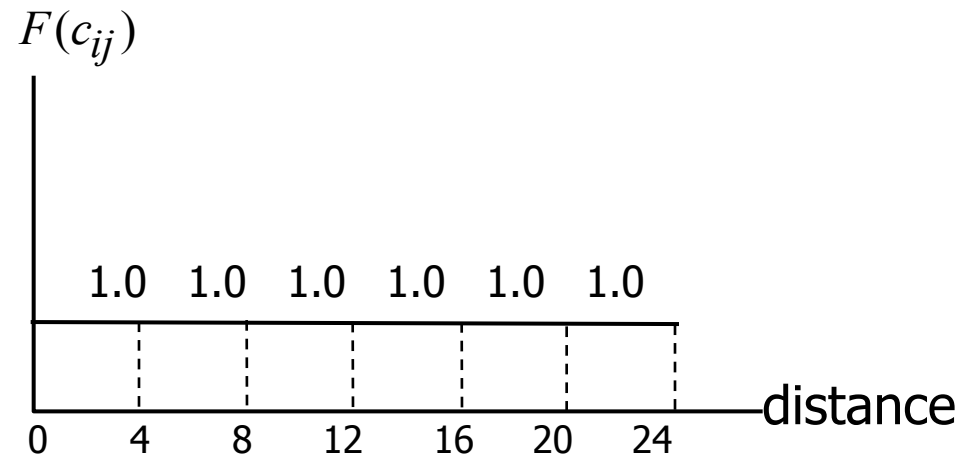
$$c_{ij} = \begin{pmatrix} 3 & 11 & 18 & 22 \\ 10 & 3 & 13 & 19 \\ 15 & 13 & 5 & 7 \\ 24 & 18 & 6 & 5 \end{pmatrix} \quad P_i = \begin{pmatrix} 400 \\ 460 \\ 400 \\ 702 \end{pmatrix}$$

$$A_j = \begin{bmatrix} 260 & 400 & 500 & 802 \end{bmatrix}$$

Example Poisson estimator: step 1

					P_i	
	1	1	1	1	400	×100
	1	1	1	1	460	×115
	1	1	1	1	400	×100
	1	1	1	1	702	×175.5
A_j	260	400	500	802		

- Start with $F(c_{ij}) = 1$
- Scale to the productions



Example Poisson estimator: step 2

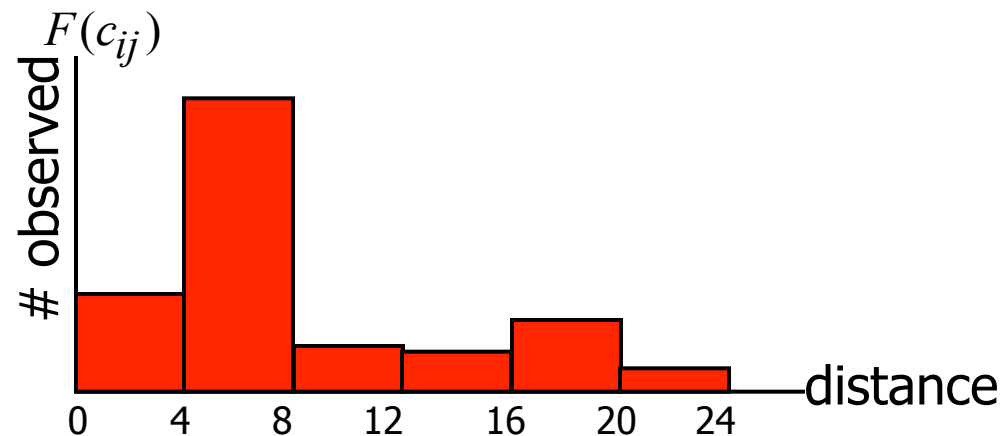
					P_i
A_j	100	100	100	100	400
	115	115	115	115	460
	100	100	100	100	400
	175.5	175.5	175.5	175.5	702
	260	400	500	802	
	×0.53	×0.82	×1.01	×1.63	

- Scale to the attractions

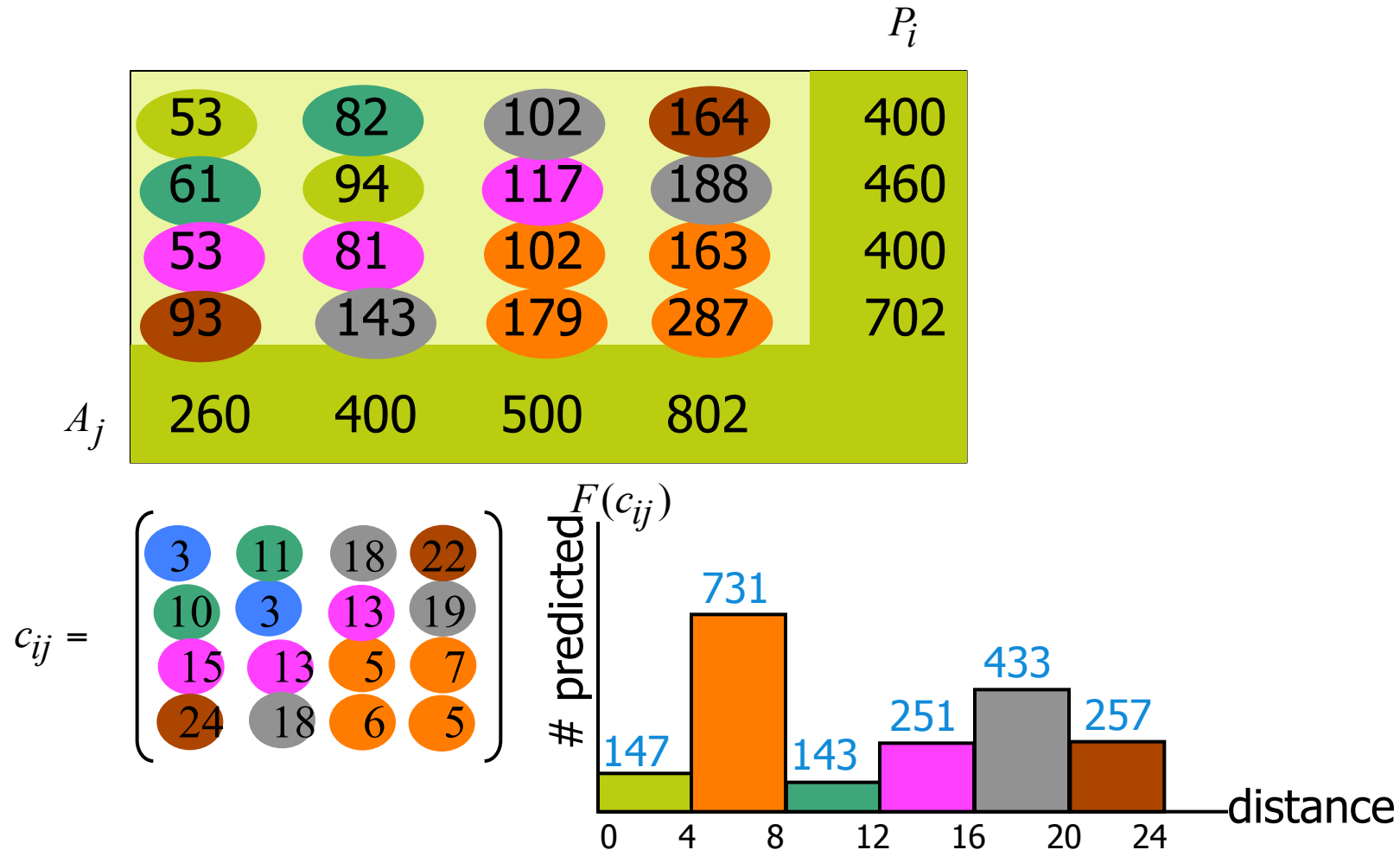
Example Poisson estimator: step 3

					P_i
A_j	53	82	102	164	400
	61	94	117	188	460
	53	81	102	163	400
	93	143	179	287	702
	260	400	500	802	

- Scale the distribution values such that they represent the OTLD



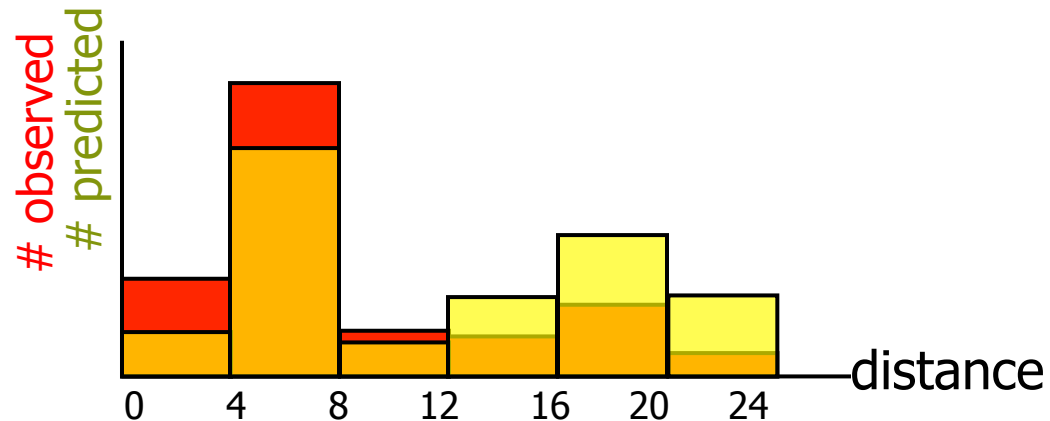
Example Poisson estimator: step 3



Example Poisson estimator: step 3

					P_i
	53	82	102	164	400
	61	94	117	188	460
	53	81	102	163	400
	93	143	179	287	702
A_j	260	400	500	802	

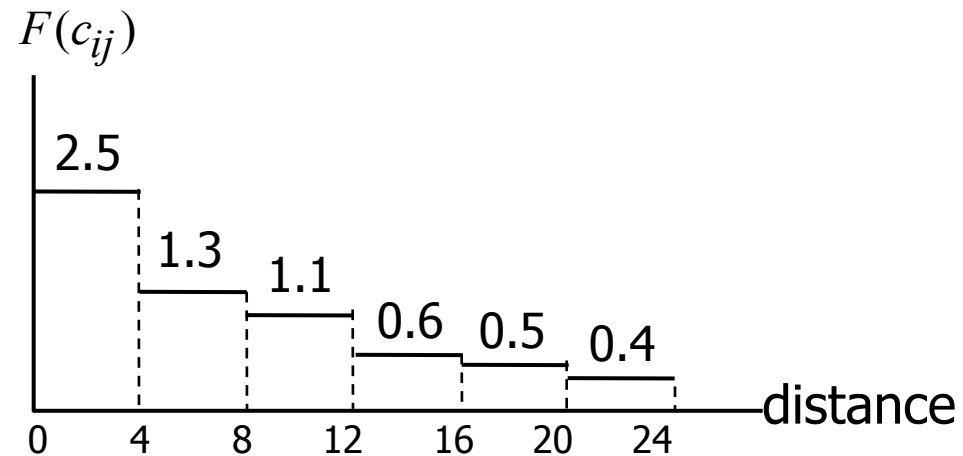
- Scale the distribution values such that they represent the OTLD



Example Poisson estimator: step 3

	P_i				
	53×2.5	82×1.1	102×0.5	164×0.4	400
	61×1.1	94×2.5	117×0.6	188×0.5	460
	53×0.6	81×0.6	102×1.3	163×1.3	400
	93×0.4	143×0.5	179×1.3	287×1.3	702
A_j	260	400	500	802	

$$c_{ij} = \begin{pmatrix} 3 & 11 & 18 & 22 \\ 10 & 3 & 13 & 19 \\ 15 & 13 & 5 & 7 \\ 24 & 18 & 6 & 5 \end{pmatrix}$$



Example Poisson estimator: result iteration 1

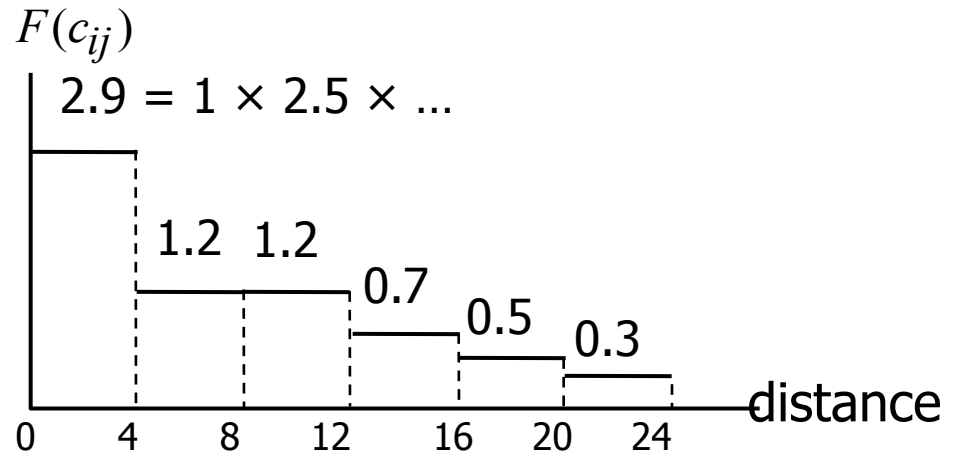
					P_i
A_j	132	92	54	61	400
	68	233	70	100	460
	32	49	134	215	400
	34	76	235	377	702
	260	400	500	802	

- Perform next iteration
 - scale to productions
 - scale to attractions
 - scale distribution values
 - etc.

Example Poisson estimator: result

					P_i	
	156	101	69	74	400	
	58	208	85	109	460	
	26	39	121	214	400	
	20	52	225	405	702	
A_j	260	400	500	802		OD-matrix after 10 iterations

Distribution
function:



From discrete distribution function to continuous function

- Just test which function yields the best fit with the function values
 - Function type and parameters
- In practice it's likely that you have to choose for which range of costs the fit is best
 - “One size doesn't fit all”

Note that the Poisson model yields estimates for Q_i and X_j as well, which can be used for a direct demand model

Alternative: Hyman's method

No OD-information available

Given:

- observed trip production
- observed trip attraction
- observed mean trip length (MTL)
- assumption: $F(c_{ij}) = \exp(-\alpha c_{ij})$, α is unknown

OD matrix can then be determined by:

using the doubly constrained gravity model
while updating α to match the MTL.

Example Hyman's method

Given:

$$c_{ij} = \begin{pmatrix} 3 & 11 & 18 & 22 \\ 12 & 3 & 13 & 19 \\ 15 & 13 & 5 & 7 \\ 24 & 18 & 8 & 5 \end{pmatrix} \quad P_i = \begin{pmatrix} 400 \\ 460 \\ 400 \\ 702 \end{pmatrix}$$
$$A_j = \begin{bmatrix} 260 & 400 & 500 & 802 \end{bmatrix}$$

Observed: MTL = 10

$$F(c_{ij}) = \exp(-\alpha c_{ij}), \quad \alpha \text{ is unknown}$$

Compute the OD-matrix and distribution function that fits the MTL.

Example Hyman's method: iteration 1

					P_i
A_j	156	99	68	77	400
	58	204	103	95	460
	26	45	138	191	400
	20	52	191	439	702
	260	400	500	802	

$$c_{ij} = \begin{pmatrix} 3 & 11 & 18 & 22 \\ 12 & 3 & 13 & 19 \\ 15 & 13 & 5 & 7 \\ 24 & 18 & 8 & 5 \end{pmatrix}$$

- First step: Choose $\alpha = 1 / \text{MTL} = 0.1$
- Compute trip distribution using a gravity model
- Compute modelled MTL

$$\text{MTL}_1 = \frac{156 \times 3 + 99 \times 11 + 68 \times 18 + 77 \times 22 + 58 \times 12 + \dots}{156 + 99 + 68 + 77 + 58 + \dots} = 8.7$$

Example Hyman's method: result

					P_i
A_j	112	98	81	109	400
	66	156	109	129	460
	39	60	120	181	400
	43	86	190	383	702
	260	400	500	802	

$$c_{ij} = \begin{pmatrix} 3 & 11 & 18 & 22 \\ 12 & 3 & 13 & 19 \\ 15 & 13 & 5 & 7 \\ 24 & 18 & 8 & 5 \end{pmatrix}$$

- Set next α
$$n = 1 : \alpha_2 = \frac{MTL_1}{MTL} \alpha_1,$$
- etc
$$n \geq 1 : \alpha_{n+1} = \frac{(MTL - MTL_{n-1})\alpha_n - (MTL - MTL_n)\alpha_{n-1}}{MTL_n - MTL_{n-1}}$$
- Choose $\alpha = 0.0586$

$$MTL = \frac{112 \times 3 + 98 \times 11 + 81 \times 18 + 109 \times 22 + 66 \times 12 + \dots}{112 + 98 + 81 + 109 + 66 + \dots} = 10.0$$

Estimation of the distribution function

Poisson model

Given:

- Cost matrix
- (partial) observed OD-matrix

Thus also known:

- (partial) production and attraction
- Totals per class for the travel costs

Estimated:

Q_{ij} , X_j and F_k

Hyman's method

Given:

- Cost matrix
- Production and attraction
- Mean trip length

Assumed:

- Type of distribution function

Estimated:

- Q_{ij} , X_j and parameter distribution function

Solution method (1 / 2)

- Basic model: $P_{ij} = Q_i \cdot X_j \cdot f(c_{ij})$
 $f(c_{ij}) = F_k, \quad c_{ij} \in \text{class}(k)$
 $f(c_{ij}) = e^{-\alpha \cdot c_{ij}}$
- With the following constraints:
 - (partial) Production and attraction
 - Totals per cost class k (Poisson)
or mean trip length (Hyman)

Solution method (2/2)

Poisson model

1. Set values for F_k equal to 1
2. Balance the (partial) matrix for the (partial) production
3. Balance the (partial) matrix for the (partial) attraction
4. Balance the (partial) matrix for the totals per cost class (i.e. Correction in iteration i for estimate of F_k)
5. Go to step 2 until convergence is achieved

$$F_k^i = \prod_j \frac{\text{observed total}_k}{\text{computed total}_k^j}$$

Hyman's method

1. Set parameter α of the distribution function equal to 1
2. Determine values for $f(c_{ij})$
3. Balance the matrix for the production
4. Balance the matrix for the attraction
5. Repeat steps 3 and 4 until convergence is achieved
6. Determine new estimate for α based on observed MTL and computed MTL and go to step 2

1.4

Building OD-matrices
Travel time functions

Travel time functions

- BPR-function is mostly used
- Different road types have different parameters
- Literature suggests some other functions based on either theoretical requirements or on computational advantages (or both)
- Focus is more on improved assignment techniques that are based on traffic flow theory

2.

Estimating matrices

What's next?

- Suppose we've got all the data and functions. So we can build an OD-matrix and assign it to the network
- Key question then is:
 - Does it match the counts?
- Common finding
 - There are quite some differences
- Two options:
 - Check the work done (especially the network!) and build a new (better) matrix
 - Adjust the OD-matrix itself to achieve a better match



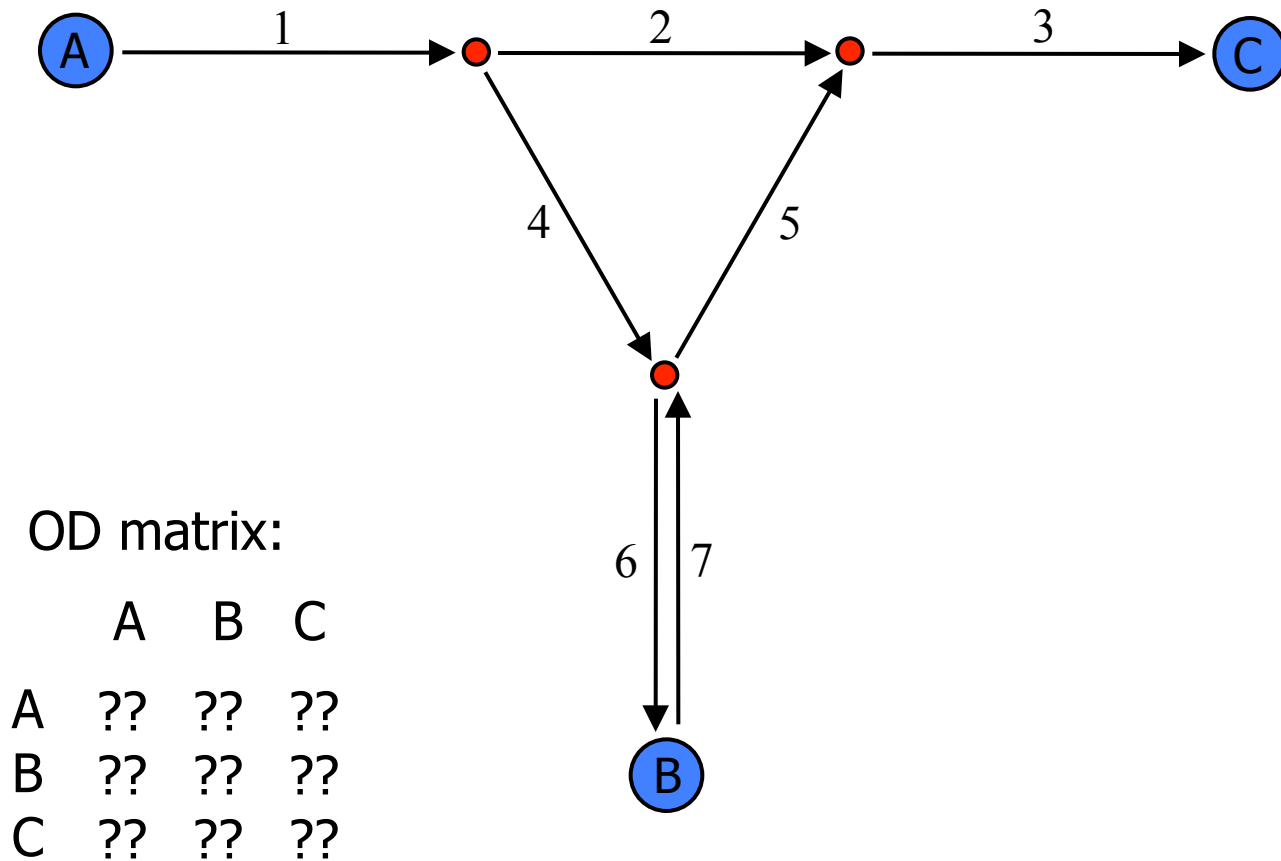
2.1

Estimating OD-matrices
Updating using traffic counts

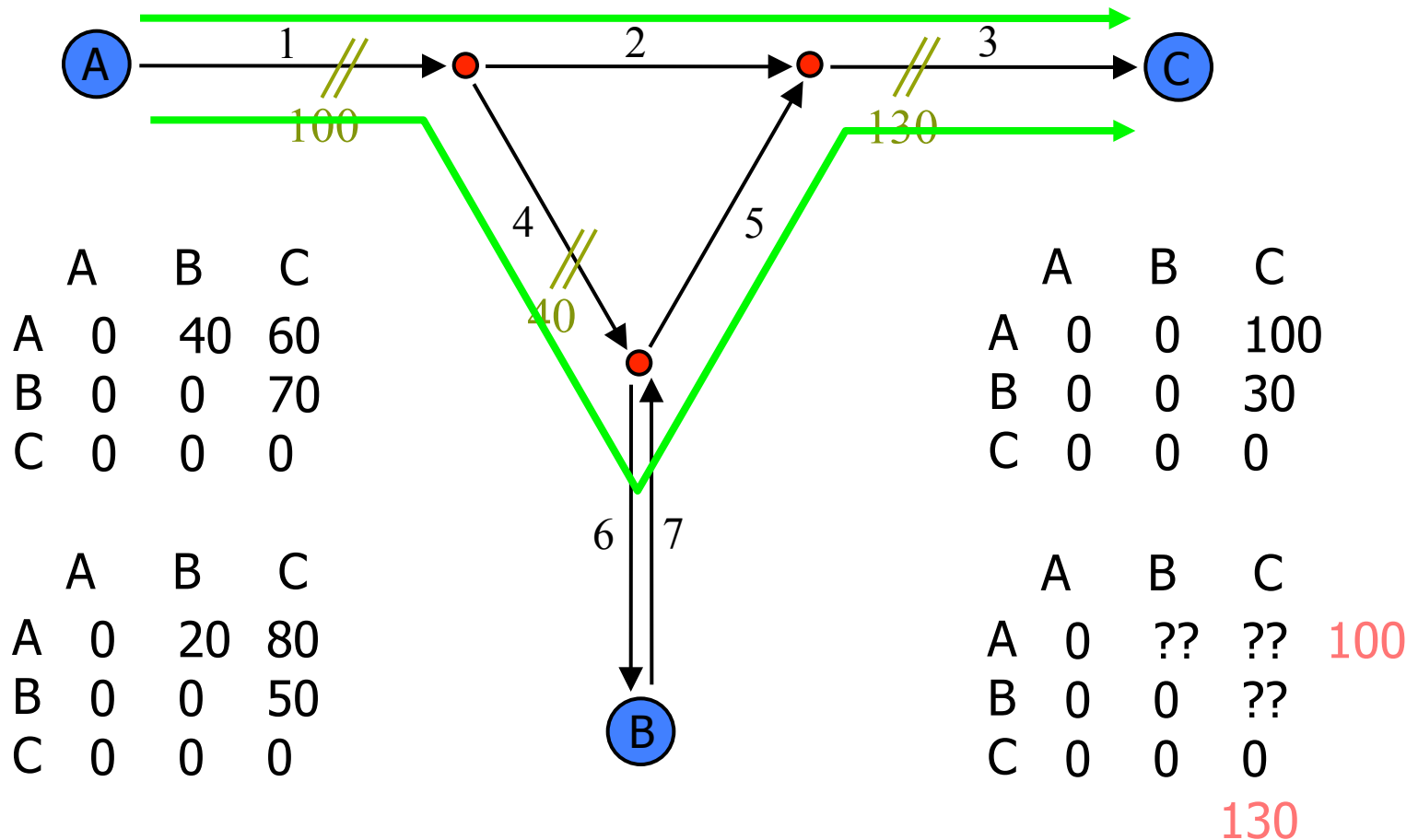
Updating an OD-matrix to traffic counts

- Counts set additional constraints for the OD-matrix
- Previously the constraints were only $\sum_j t_{ij} = P_i$ and $\sum_i t_{ij} = A_j$
- The constraint for a count on link a is $\sum_i \sum_j \sum_r \alpha_{ijr}^a \cdot \beta_{ijr} \cdot t_{ij} = S_a$
- So, one option is to start with a simple matrix ($t_{ij}=1$) and balance the matrix to fit all constraints
 - Note that the distribution function and possible segmentations are omitted
 - Furthermore, this only works if all constraints are consistent.....
- The second option is to find a matrix that finds a balance between the original (a priori matrix) and the counts=> optimisation

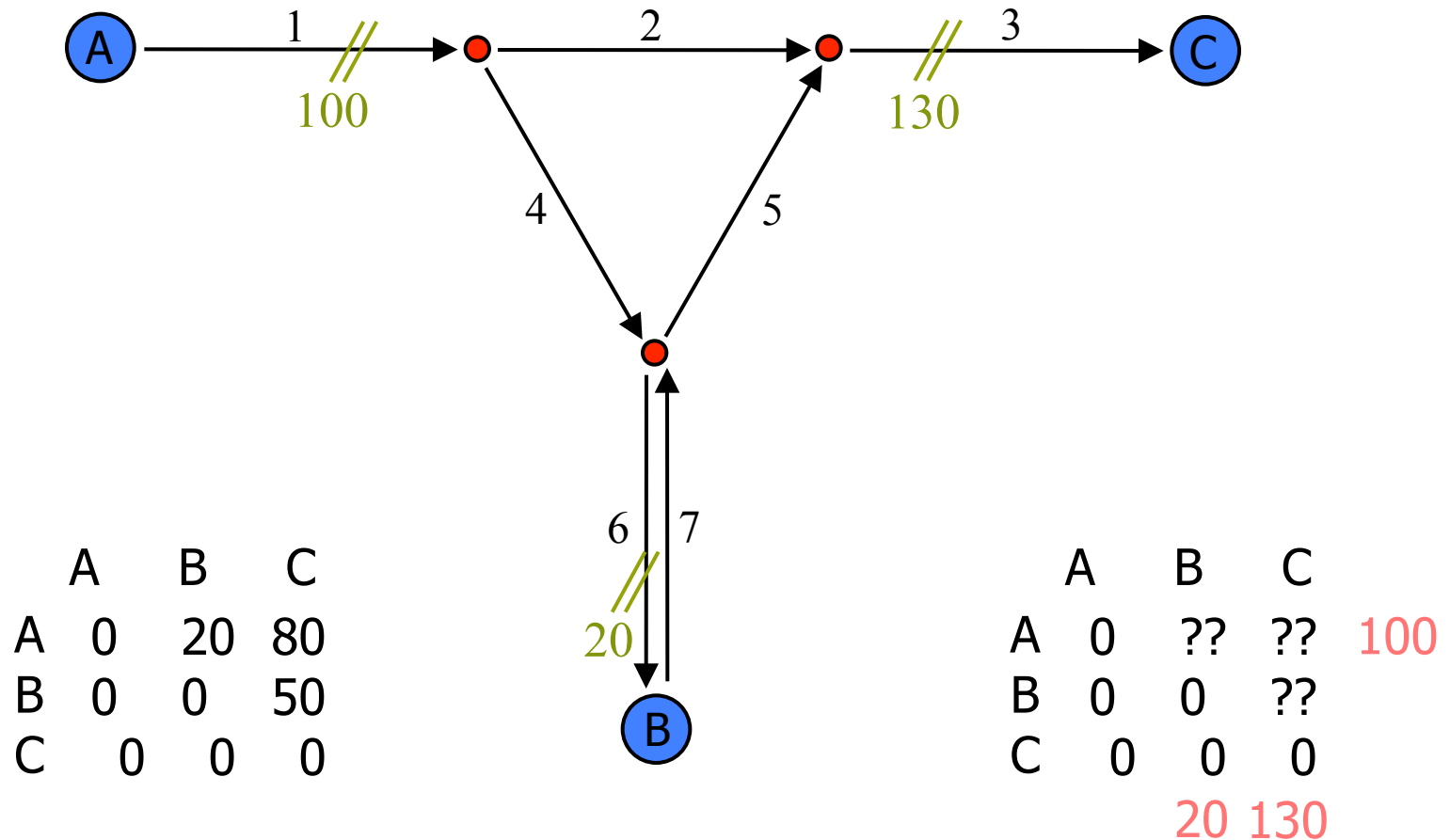
Calibration from link counts only



Calibration from link counts only



Calibration from link counts only



Alternative approach

- Use an a priori matrix from a model and assign it to the network
- For each link having a count the ideal multiplier for the OD-pairs using that link is

$$x_k = \frac{S_k}{\sum_i \sum_j \sum_r \alpha_{ijr}^k \cdot \beta_{ijr} \cdot t_{ij}}$$

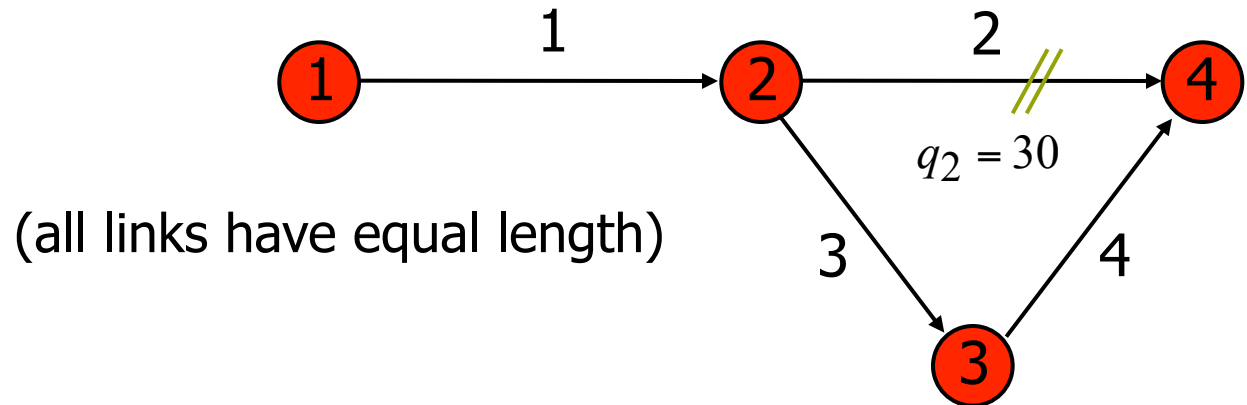
- As an OD-pair might pass multiple counts the resulting multiplier for that OD-pair is

$$y_{ij} = \frac{\sum_k \left(x_k \cdot \sum_r \alpha_{ijr}^k \cdot \beta_{ijr} \cdot t_{ij} \right)}{\sum_k \sum_r \alpha_{ijr}^k \cdot \beta_{ijr} \cdot t_{ij}}$$

i.e. the weighted average of the ideal multipliers for a specific OD-pair

- Obviously, the averaging of the ideal multipliers per count leads to a poorer match than intended, thus (again) an iterative procedure

Example based on AON



OD-matrix

0	10	15	20
0	0	10	5
0	0	0	20
0	0	0	0

$$\frac{30}{20+5} \cdot 20 = 24$$

$$\frac{30}{20+5} \cdot 5 = 6$$

Quality of constraints

- Not all counts have the same quality, e.g. permanent loop detectors versus manual counts between 07.00 and 19.00
- Trick: assign a weight (or “elasticity”) e_k to each constraint ranging between 0 (no influence at all) and 1 (maximum adjustment)
- The multiplier for each constraint then becomes

$$x_k = \left(\frac{S_k}{\sum_i \sum_j \sum_r \alpha_{ijr}^k \cdot \beta_{ijr} \cdot t_{ij}} \right)^{e_k}$$

- Consequence: not all constraints are met at all costs

2.2

Estimating OD-matrices Formulation as optimisation problem

Criteria for differences between OD-matrices

- Least squares between estimated number of trips and the a priori number of trips

$$\sum_{ij} \left(\hat{T}_{ij} - T_{ij}^{ap} \right)^2$$

- Minimum distance criterium based on information minimisation

$$\sum_{ij} \left(\hat{T}_{ij} \cdot \ln \left(\frac{\hat{T}_{ij}}{T_{ij}^{ap}} \right) - \hat{T}_{ij} + T_{ij}^{ap} \right)$$

Formulation as optimisation problem

- Given the objective function (criterion) the constraints can be included using Lagrange multipliers, e.g.

$$L = \sum_{ij} \left(\hat{T}_{ij} \cdot \ln \left(\frac{\hat{T}_{ij}}{T_{ij}^{ap}} \right) - \hat{T}_{ij} + T_{ij}^{ap} \right) - \sum_k \lambda_k \cdot \left(S_k - \sum_i \sum_j \sum_r \alpha_{ijr}^k \cdot \beta_{ijr} \cdot \hat{T}_{ij} \right)$$

- Setting the derivatives equal to zero yields a set of equations that can be solved using dedicated software

Two comments

- Note that this formulation can also be used for other types of constraints such as production, attraction or classes of observed trip length distribution
- These methods can also deal with probability distributions for the constraints
 - Including covariance between observations

3.

Special topics

Special topics

- Ratio between variables and constraints?
- Which adjustments are acceptable?
- What about OD-pairs that are not affected by constraints?
- Matrix estimation in congested networks?

Ratio between variables and constraints

- Number of variables: n^2
- Number of constraints:
 - Production and attraction: $2n$
 - Counts: depends on the size of model
 - Municipality: a few dozen
 - Region: a few hundred
 - In order to reduce impacts of "route choice", counts are often aggregated into screenlines => less constraints
- Problem is heavily underspecified

Which adjustments are acceptable?

- In many cases there are OD-pairs that relate to a single count
- Consequence is that a perfect match is always possible
- However, is this really perfect?

What about OD-pairs that are not affected by constraints?

- Suppose that 60% of the trips from a zone pass a count
- Suppose that these 60% show an overall increase of 10%
- What is then realistic for the other 40% of the trips from that zone:
 - Should be decreased to match the original production?
 - Should be increased by 10% as well?

Matrix estimation in congested networks?

- In OD-estimation you try to match the modelled flow with counts
- In congested networks the flows are related to capacities
- In fact, the assignment already tries to limit the flows to the capacities
- So what is the added value of matrix estimation?
- Furthermore: if the OD-matrix changes the assignment changes as well.....