Characterizing Travel Time Reliability and Passenger Path Choice in a Metro Network

Lijun SUN

Future Cities Laboratory, Singapore-ETH Centre

lijun.sun@ivt.baug.ethz.ch

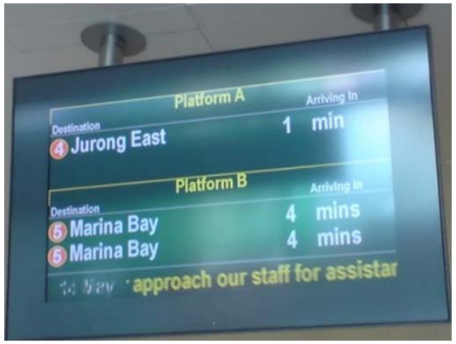
National University of Singapore (NUS)

(FCL) FUTURE CITIES LABORATORY 未来城市实验室

(SEC) SINGAPORE-ETH CENTRE 新加坡－ETH研究中心

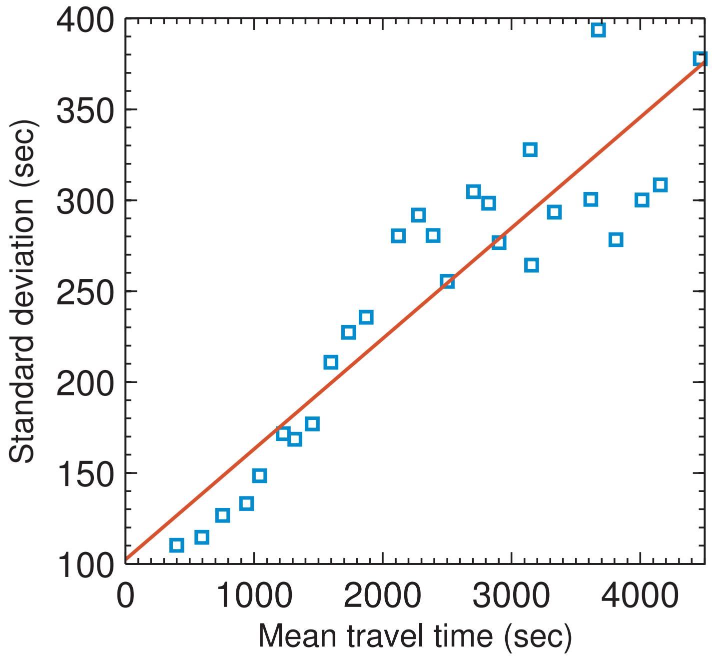
Background

- Service Reliability
- Travel time
- Train load
- Transfer convenience
- Service disruption
- Flow assignment



Background

- Service Reliability
- The reliability of metro systems is higher than other transit modes. However, travel time variability still shows accumulative effect.



Background

- Most metro systems are closed environments, which only register transactions when passengers enter and leave the system

- Crucial questions:
- Predicting travel time (and reliability)
- Inferring route choices
- Inferring train load
- Identifying critical transfer location
- Building sophisticated flow assignment models
- All linked together……

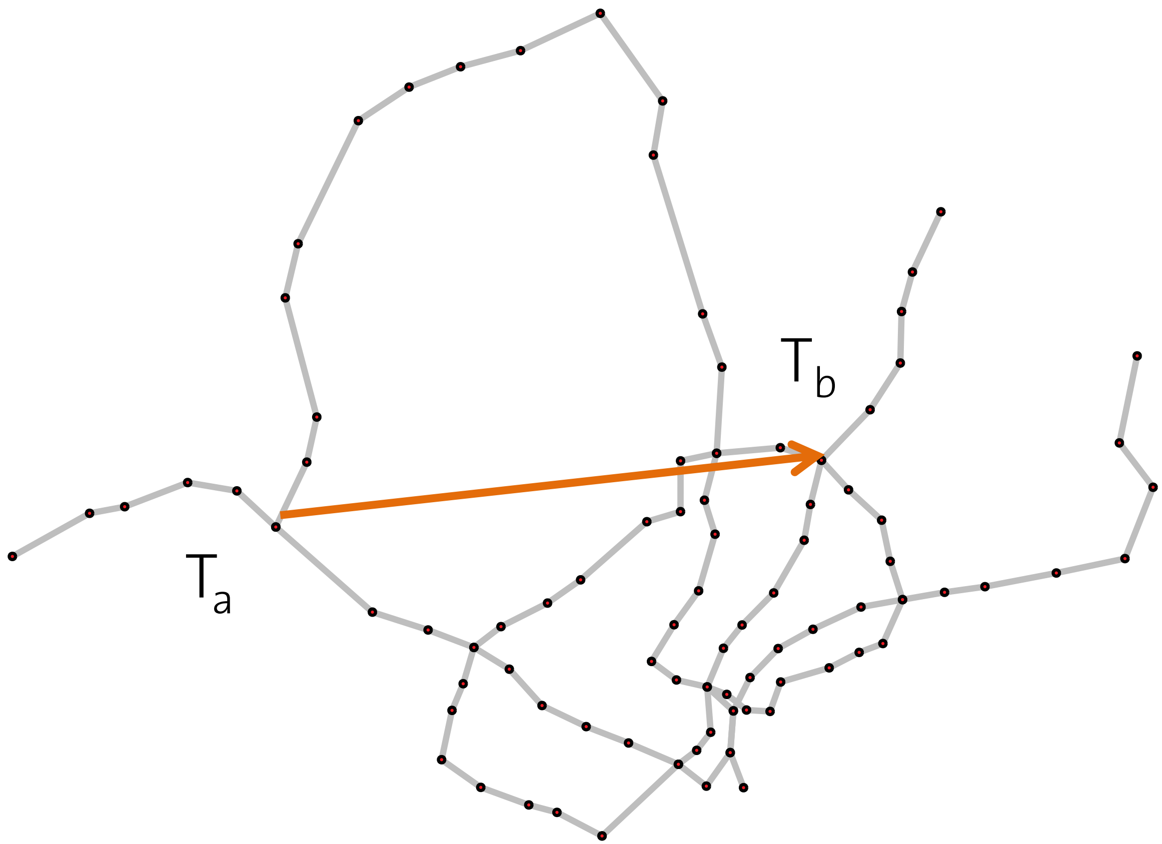
Background

- **Data**

- Operation log?
- Transfer demand?
- Link flow?
- Route choice?
- Trajectory?

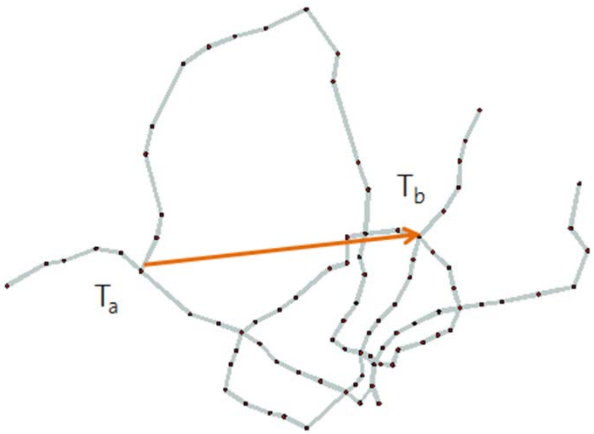
- Smart card (Boarding station, Alighting station, Travel time)

Metro System



Metro System

- What we know? (Observed)
  - Network configuration
  - Boarding station B
  - Alighting station A
  - Travel time t



- What we do not know? (Unknown)
  - Travel time on each link
  - Reliability of each link
  - Other time cost (waiting at platform, walking between fare gantry and platform)
  - Route choice

Metro System

- Research question
- Given observations to infer unknowns

observed

- Boarding station B
- Alighting station A
- Travel time t

unknown

- Travel time on each link
- Reliability of each link
- Other time cost
- Route choice

Metro System

- **Methodology: Bayesian inference**
- **P(unknown|observed) ∝ P(observed|unknown) x P(unknown)**
- **Likelihood x Prior**

observed

- Boarding station B
- Alighting station A
- Travel time t

unknown

- Travel time on each link
- Reliability of each link
- Other time cost
- Route choice

Unknown

- Travel time on each link
- Travel time variation on each link (in-vehicle / transfer links)
  - Assuming that link cost follows a normal distribution, which has a constant coefficient of variation (linear mean v.s. std)

$$x_a \sim \mathcal{N}\left(c_a, (\alpha c_a)^2\right)$$

  - Assuming that all links are independent
  - Then the travel time on a particular route r follows

$$t \mid r \sim \mathcal{N}\left(\sum_{a\in r} c_a, \alpha^2 \sum_{a\in r} c_a^2\right)$$

Unknown

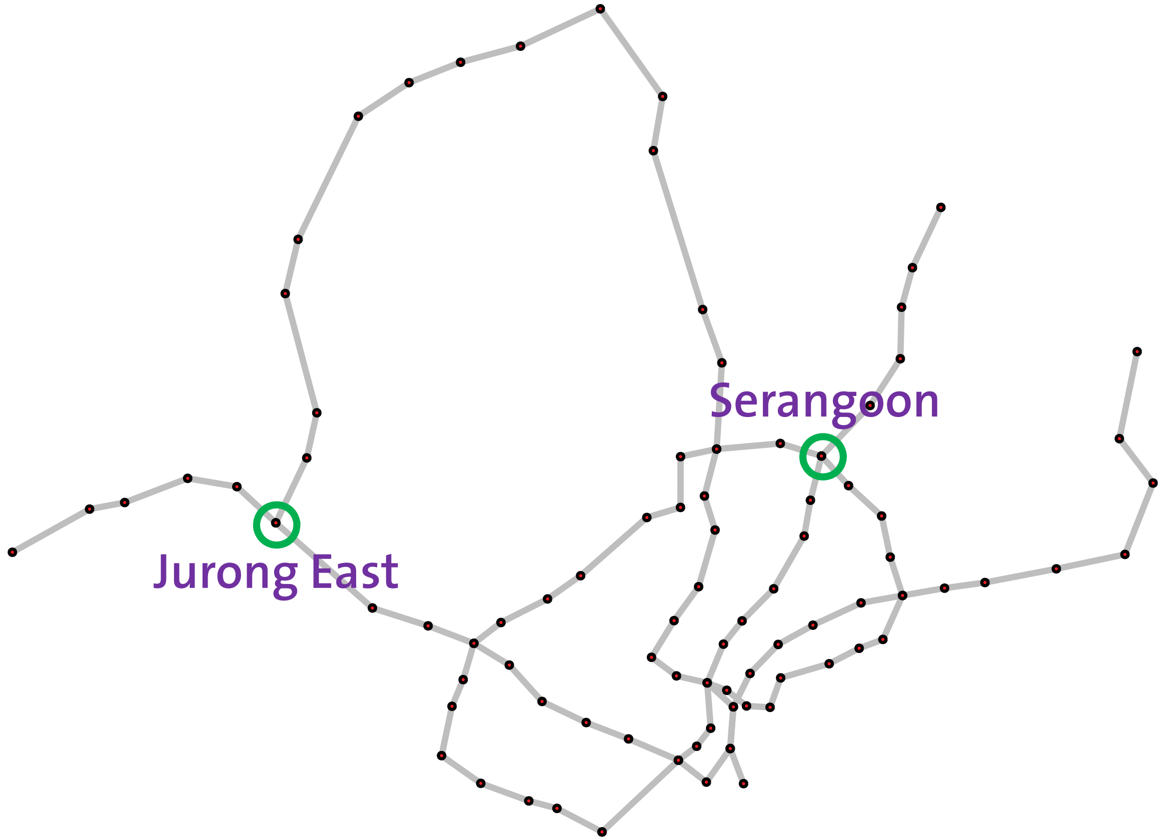
- Other time cost
  - Assuming that other cost follows a normal distribution, with

$$y \sim \mathcal{N}\left(m, \sigma_y^2\right)$$

  - Assumed to be consistent for all OD pairs

Unknown

- Route choice



Unknown

- Which route to take?
- Using brute-force search

Unknown

- Which route to take?
- Using brute-force search

Unknown

- Which route to take?
- Using brute-force search

Unknown

- Which route to take?
- Using brute-force search

Unknown

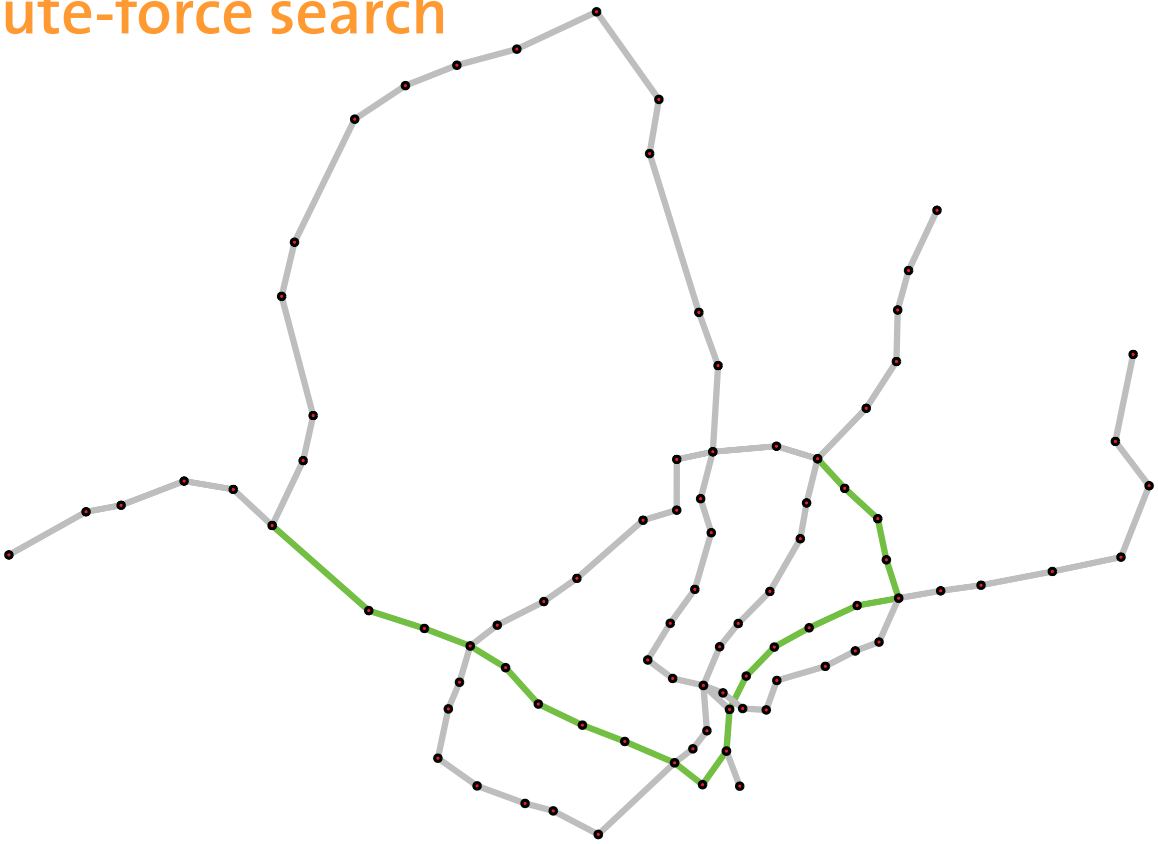
- Which route to take?
- Using brute-force search

Unknown

- Which route to take?
- Using brute-force search

Unknown

- Which route to take?
- Using brute-force search

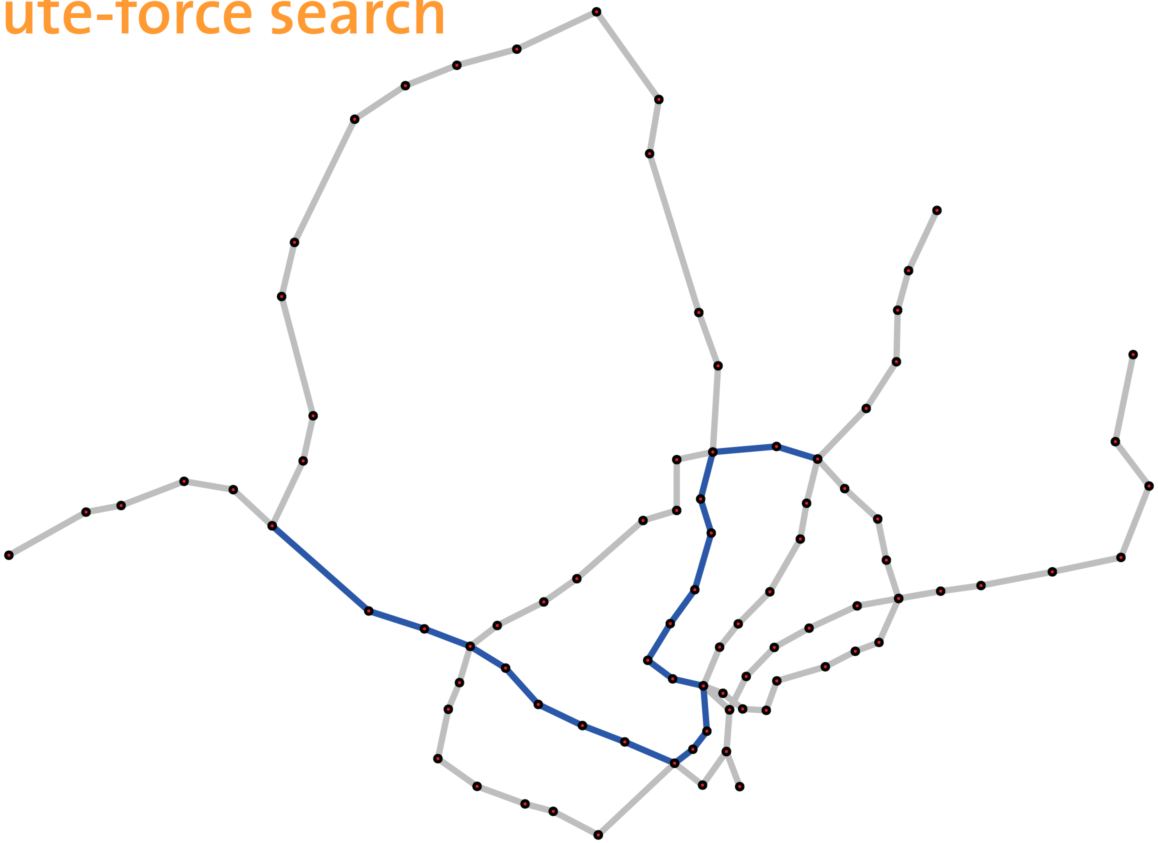


Unknown

- Which route to take?
- Using brute-force search

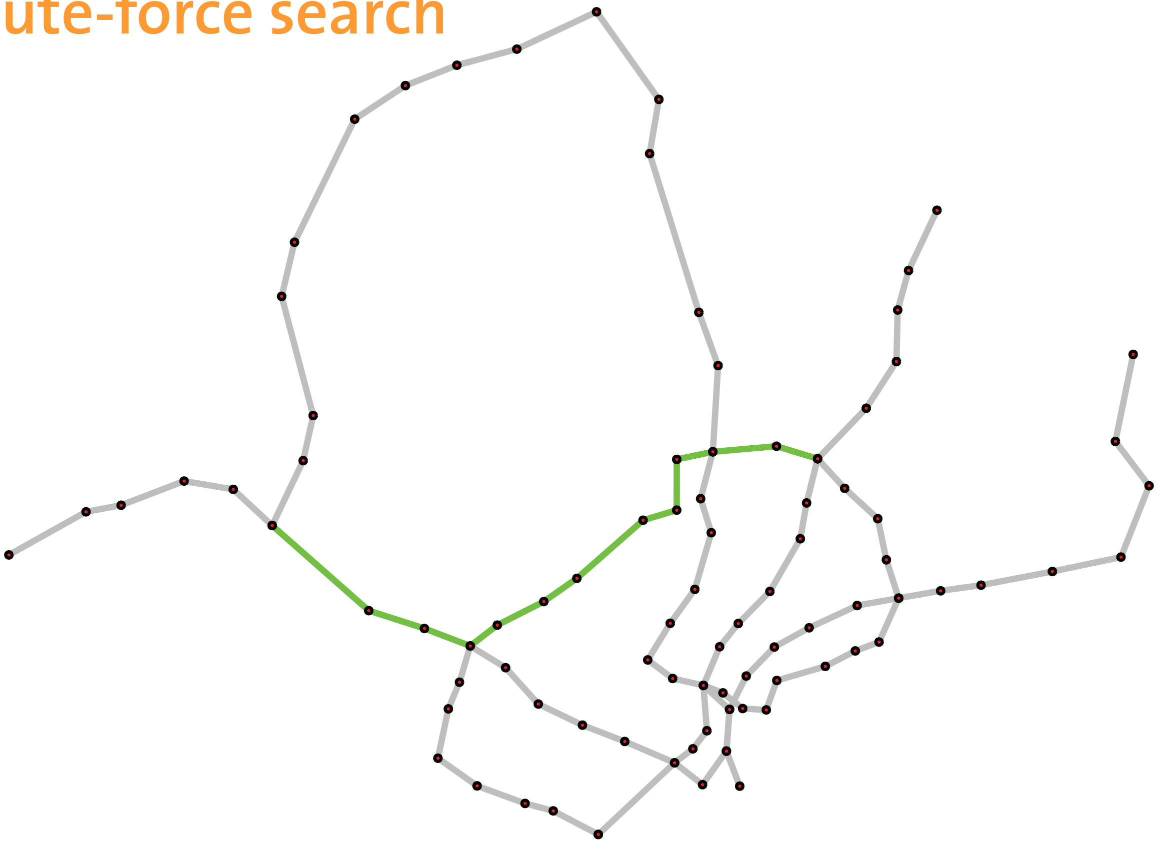
Unknown

- Which route to take?
- Using brute-force search



Unknown

- Which route to take?
- Using brute-force search



Unknown

- Which route to take?
- Using brute-force search

More than 30 routes in total

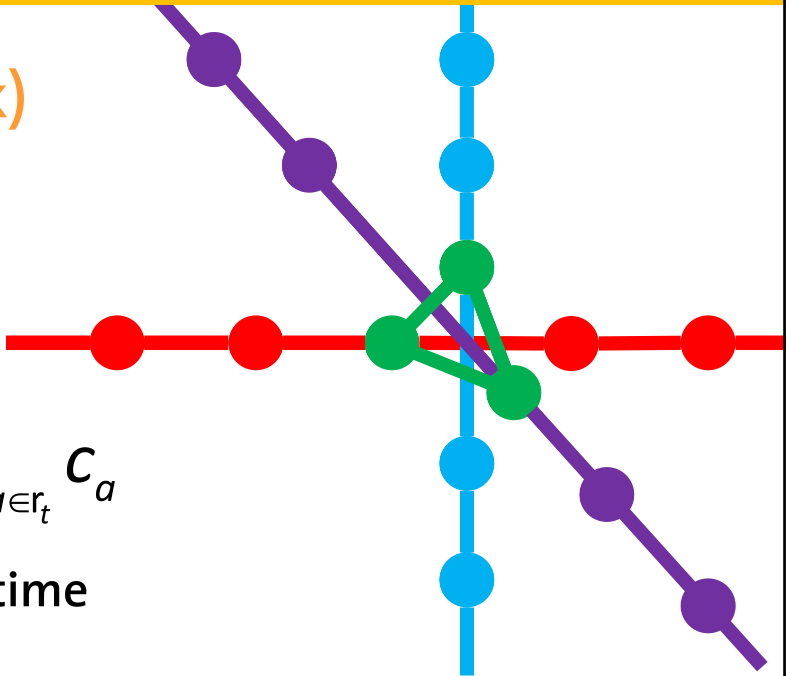
Unknown

- Route choice (in general)
- Multinomial Logit (MNL)
- Representative Utility $V_i = \sum_k \theta_k X_{ik}$
- Choice probability $P_i = \dfrac{\exp(V_i)}{\sum_{j \in A} \exp(V_j)}$
- Parameters θ_k are unknown

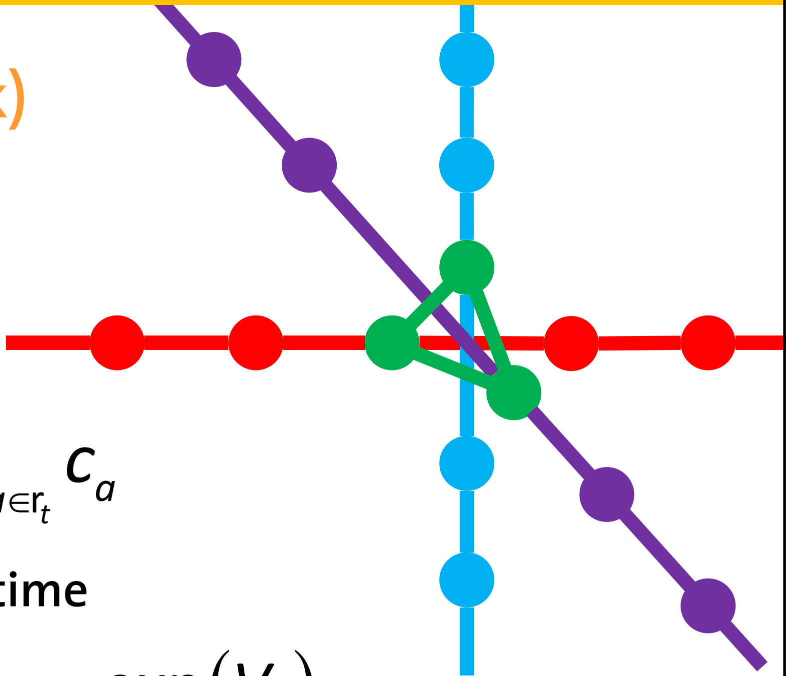
Unknown

- Route choice (for the metro network)
- Multinomial Logit (MNL)
- Utility $V_r = \theta_1 \times \sum_{a \in r \setminus r_t} c_a + \theta_2 \times \sum_{a \in r_t} c_a$

in-vehicle time　　transfer time



Unknown



- Route choice (for the metro network)
- Multinomial Logit (MNL)
- Utility $V_r = \theta_1 \times \sum_{a \in r \setminus r_t} c_a + \theta_2 \times \sum_{a \in r_t} c_a$

  in-vehicle time — transfer time
- Choice probability $f_w(r \mid \mathbf{c}, \alpha, \boldsymbol{\theta}) = \frac{\exp(V_r)}{\sum_{r \in R_w} \exp(V_r)}$

$|R_w| \geq 1$

For each OD pair w

Observed

- Observing travel time $t = t_b - t_a$ for OD pair $w = (a, b)$
- The probability observing t on route r

$$t \mid r \sim \mathcal{N}\left(\sum_{a \in r} c_a + m, \alpha^2 \sum_{a \in r} c_a^2 + \sigma_y^2 \right)$$

- The probability observing t on OD pair w

$$p_w\left(t \mid \mathbf{c}, \alpha, \boldsymbol{\theta}, m\right) = \sum_{r \in R_w} h\left(t \mid r\right) f_w\left(r \mid \mathbf{c}, \alpha, \boldsymbol{\theta}, m\right)$$

Observed

- The likelihood of observation all smart card transactions
- (travel time)

$$\mathcal{L}\left(\mathbf{c},\alpha,\boldsymbol{\theta},m \mid \mathbf{T}\right)=\prod_{w\in W} p\left(\mathbf{T}_w \mid \mathbf{c},\alpha,\boldsymbol{\theta},m\right)$$

$$=\prod_{w\in W}\left(\prod_{t\in \mathbf{T}_w}\left(\sum_{r\in R_w} h\left(t\mid r\right) f_w\left(r\mid \mathbf{c},\alpha,\boldsymbol{\theta},m\right)\right)\right)$$

Observed

- Prior knowledge
- Mean link travel time follows normal distribution

$$c_a \sim \mathcal{N}(2,1)$$

  - Travel time between stations / transfer time: around 2 minutes

- Other cost follows a normal distribution

$$m \sim \mathcal{N}(4,1)$$

  - Waiting time plus access/egress cost: around 4 minutes in total

Observed

- **Prior knowledge**
- **Parameters for MNL: we do not have any information**
- **In the literature:**

- Raveau S, Guo Z, Muñoz JC, & Wilson NHM (2014) A behavioural comparison of route choice on metro networks: Time, transfers, crowding, topology and socio-demographics. *Transportation Research Part A: Policy and Practice* 66:185-195.

Table 2
Parameters estimates.

Attribute	London Underground		Santiago Metro	
	Parameter	*t*-Value	Parameter	*t*-Value
In-vehicle time	−0.121	−9.20	−0.074	−6.30
+Morning peak	−0.084	−5.10	−0.014	−2.52
+Afternoon peak	n.a.[a]	n.a.	−0.009	−2.62
+Restrictive Purpose	−0.042	−2.42	−0.025	−5.93
Waiting time	−0.269	−14.21	−0.083	−3.62
+Morning peak	−0.208	−5.94	−0.094	−2.60
Walking time	−0.299	−9.32	−0.210	−2.34
+Women	−0.048	−2.06	−0.074	−2.67
Number of transfers	−1.321	−4.14	−0.662	−4.19
Ascending transfers	−0.206	−2.53	−0.308	−2.60
Even transfers	0.613	3.82	n.a.[b]	n.a.
Descending transfers	0.000 [c]	n.a.	0.000[c]	n.a.
Assisted transfers	0.000 [c]	n.a.	0.000[c]	n.a.
Semi-assisted transfers	−0.271	−5.30	n.a.[b]	n.a.
Non-assisted transfers	−0.398	−6.33	−0.182	−5.11
Mean occupancy	−2.898	−3.25	−0.935	−5.10
Getting a seat	0.117	2.22	0.105	3.68
Not boarding	−0.502	−6.23	−0.358	−2.29
Angular cost	−0.088	−3.89	−0.029	−3.84
+Restrictive purpose	0.049	3.79	0.011	2.70
Map distance	−0.364	−5.43	−0.278	−4.83
Number of stations	−0.424	−5.07	−0.168	−3.62
Turning back	−0.650	−8.85	−0.142	−8.10
Turning away	−0.943	−7.77	−0.231	−8.87
Commonality factor	−0.396	−3.74	−0.541	−3.41
Sample size	17,073		28,961	
Log-likelihood	−6690		−12,881	
Corrected ρ^2	0.567		0.383	

Observed

- Prior knowledge
- Parameters for MNL: we do not have any information
  - We take uniform priors

$$\boldsymbol{\theta} \sim \mathcal{U}(-4,0)$$

- Coefficient of variation
  - We take a uniform prior

$$\alpha \sim \mathcal{U}(0,1)$$

Posterior

Likelihood x Prior

$$\pi(\mathbf{c},\alpha,\boldsymbol{\theta},m\,|\,\mathbf{T})$$

$$\propto \prod_{w\in W}\left(\prod_{t\in\mathbf{T}_w}\left(\sum_{r\in R_w} p(t\,|\,r)\,f_w(r\,|\,\mathbf{c},\boldsymbol{\theta})\right)\right)\times\prod_{c\in\mathbf{c}}\phi(c;2,1)\times\phi(m;4,1)$$

Solution Algorithm

- **MCMC (Markov Chain Monte Carlo)**
- **Variable-at-a-time Metropolis sampling scheme**

$$\boldsymbol{\delta} = \left(c_1, \cdots, c_N, \alpha, \theta_1, \theta_2, m\right) = \left(\delta_1, \cdots, \delta_{N+4}\right)$$

Solution Algorithm

- MCMC (Markov Chain Monte Carlo)
- Variable-at-a-time Metropolis sampling scheme

$$\boldsymbol{\delta} = \left(c_1, \cdots, c_N, \alpha, \theta_1, \theta_2, m\right) = \left(\delta_1, \cdots, \delta_{N+4}\right)$$

- STEP (0)
- Specify initial sample

$$\boldsymbol{\delta}^{(0)} = \left(c_1^{(0)}, \cdots, c_N^{(0)}, \alpha^{(0)}, \theta_1^{(0)}, \theta_2^{(0)}, m^{(0)}\right)$$

- Set t=0

Solution Algorithm

- STEP (1)
- At step t, sample in turn $\delta_i^{(t)}$ (for i = 1: N+4)
- Calculate

$$\mathcal{A}\left(\delta_i^*,\delta_i^{(t)}\right)=\min\left\{1,\frac{p\left(\mathbf{T}\mid\delta_i^*,\delta_{-i}^{(t)}\right)\pi\left(\delta_i^*,\delta_{-i}^{(t)}\right)}{p\left(\mathbf{T}\mid\delta_i^{(t)},\delta_{-i}^{(t)}\right)\pi\left(\delta_i^{(t)},\delta_{-i}^{(t)}\right)}\right\}$$

- Where $\delta_{-i}^{(t)}=\left(\delta_1^{(t+1)},\cdots,\delta_{i-1}^{(t+1)},\delta_{i+1}^{(t)},\cdots,\delta_{N+4}^{(t)}\right)$
- is the latest updated variables except $\delta_i^{(t)}$

- Accept $\delta_i^{(t+1)}=\delta_i^*$ with probability $\mathcal{A}\left(\delta_i^*,\delta_i^{(t)}\right)$
- Otherwise, set $\delta_i^{(t+1)}=\delta_i^{(t)}$

Solution Algorithm

- STEP (2)
- If $t < T$: set $t = t + 1$
  - Go to STEP (1)

- Else:
  - Stop iteration

Numerical Example: MRT Network in Singapore

- MCMC provides distribution of unknown parameters rather than one value

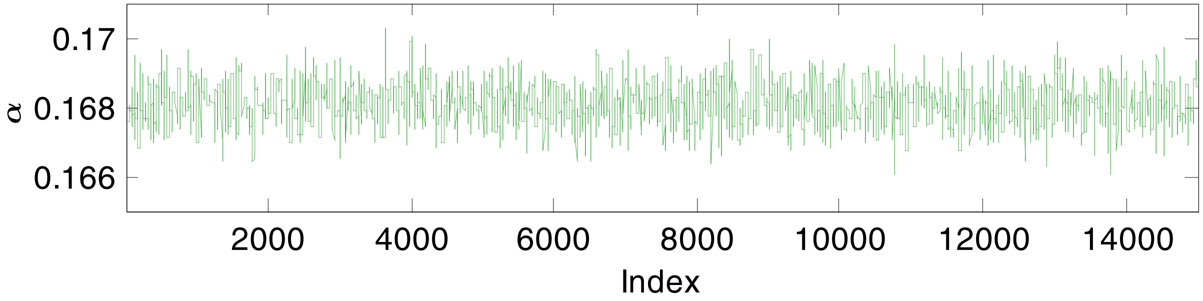
- Burn-in: 5000 steps
- Effective sample: 25000 (25000+5000 draws in total)

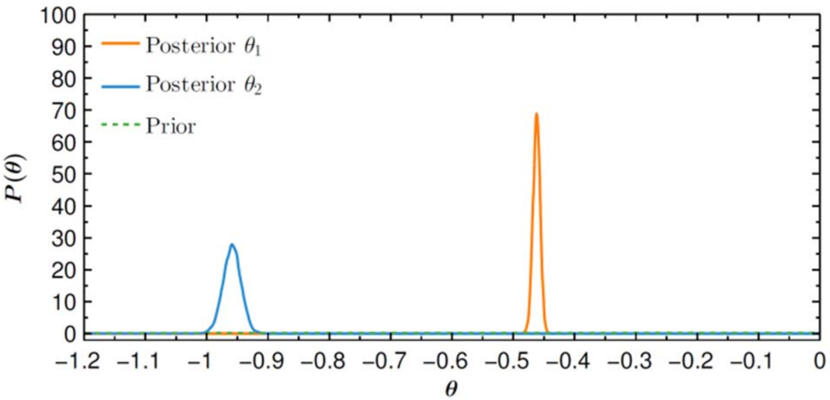
- Standard deviation for
- Gaussian random walk Metropolis proposals

$$\delta_i^* = \delta_i^{(t)} + \epsilon_i \quad \epsilon_i \sim \mathcal{N}\left(0, \sigma_i^2\right)$$

Numerical Example

- MCMC provides distribution of unknown parameters rather than one value
- α 30.8%
- θ_1 36.7%
- θ_2 45.8%

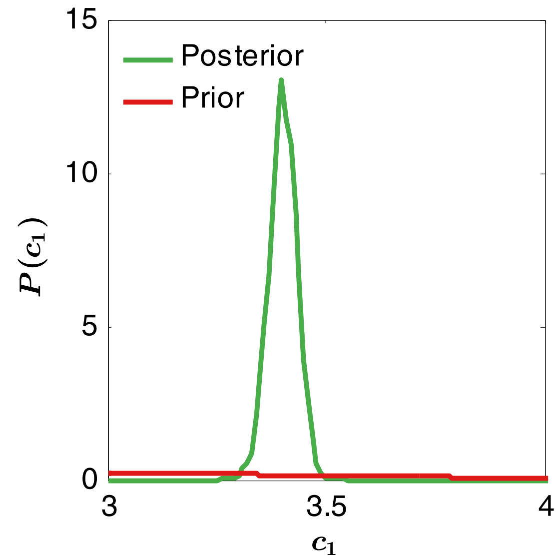




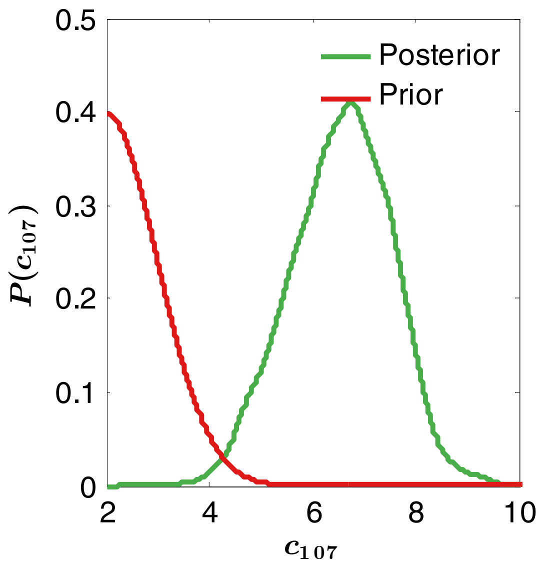
Numerical Example

- In this example

- Our prior knowledge is inaccurate
- the large number of travel time observations has corrected it



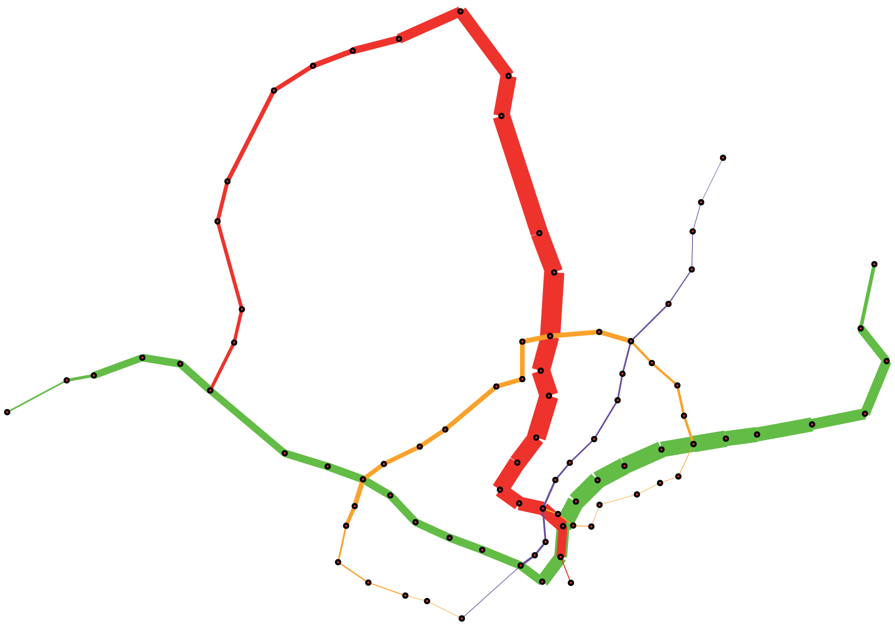
Travel time from EW1 to EW2

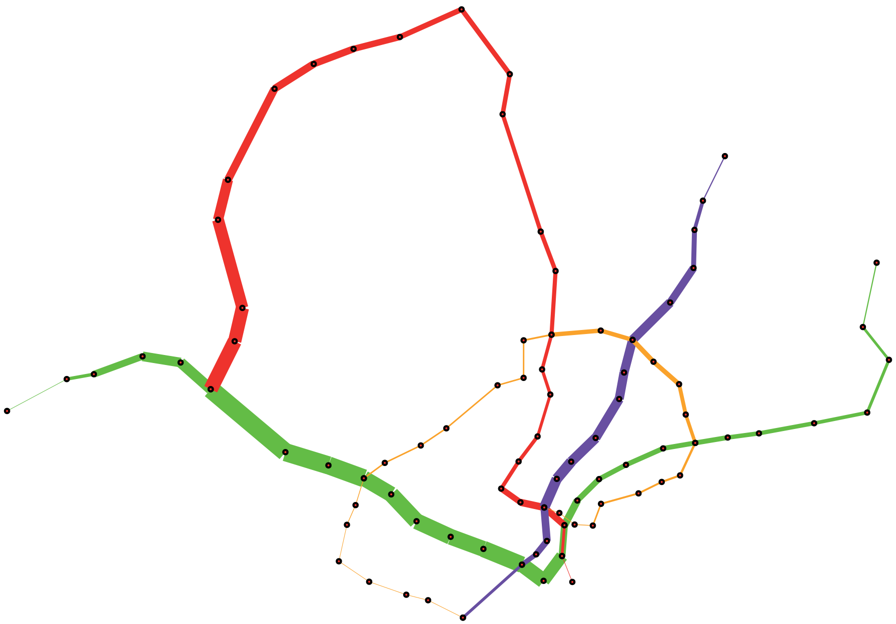


Transfer time @ Bouna Vista Stn

Numerical Example

- **Flow assignment based on route choice model**
-





Direction 1

Direction 2

Conclusion

- An integrated statistical model on travel time reliability and route choice behavior
- A metro network in which only travel time is observed
- Bayesian inference framework to formulate posterior probability
- Given the high-dimension of parameters, variable-at-a-time Metropolis sampling algorithm is applied to obtain posterior distribution.

Conclusion

- An integrated statistical model on travel time reliability and route choice behavior
- A metro network in which only travel time is observed
- Bayesian inference framework to formulate posterior probability
- Given the high-dimension of parameters, variable-at-a-time Metropolis sampling algorithm is applied to obtain posterior distribution.
- With this framework, we characterized travel time and its variation on each link. Meanwhile, we also identified contribution of different factors in determining passenger route choice behavior/movement.

Discussion & Outlook

- Most metro systems are closed environments, which only register transactions when passengers enter and leave the system; as a result, route choice (interchange/transfer) and service reliability are not captured in smart card data
- Our framework does not require an specific route choice model; thus, it can be applied on a more sophisticated model which takes more factors into account, helping us to further understand passenger behavior and build advanced flow assignment models, and further infer individual train load
- Although Singapore's network is simple, this framework shows great potential in applying on more complex metro networks, such as London Underground
- Identifying critical/crowding location/facility in metro network

Thank you!

Lijun SUN

Future Cities Laboratory, Singapore-ETH Centre

lijun.sun@ivt.baug.ethz.ch