Modeling transportation networks

Route choice models

Michel Bierlaire

TRANSP-OR Transport and Mobility Laboratory
Ecole Polytechnique Fédérale de Lausanne
tranp-or.epfl.ch
Introduction

Analysis of the supply and demand interactions

- supply: infrastructure, vehicles
- demand: travellers’ decisions
  - origin and destination
  - transportation mode
  - departure time
  - route, itinerary
- Underlying mathematical structure: the network

In this lecture, we focus on one of the most complicated issue
Introduction

Analysis of the supply and demand interactions

- supply: infrastructure, vehicles
- demand: travellers’ decisions
  - origin and destination
  - transportation mode
  - departure time
  - route, itinerary
- Underlying mathematical structure: the network

In this lecture, we focus on one of the most complicated issue
Route choice model

Given

- a mono- or multi-modal transportation network (nodes, links, origin, destination)
- an origin-destination pair
- link and path attributes

identify the route that a traveler would select.
Choice model

Assumptions about

1. the decision-maker: \( n \)

2. the alternatives
   - Choice set \( C_n \)
   - \( p \in C_n \) is composed of a list of links \( (i, j) \)

3. the attributes
   - link-additive: length, travel time, etc.
   - non link-additive: scenic path, usual path, etc.

4. the decision-rules: \( \Pr(p|C_n) \)
Shortest path

Decision-makers all identical

Alternatives

• all paths between O and D
• \( C_n = \mathcal{U} \ \forall n \)
• \( \mathcal{U} \) can be unbounded when loops are present

Attributes one link additive generalized cost

\[
c_p = \sum_{(i,j) \in P} c(i,j)
\]

• traveler independent
• link cost may be negative
• no loop with negative cost must be present so that \( c_p > -\infty \) for all \( p \)
Shortest path

**Decision-rules** path with the minimum cost is selected

\[
Pr(p) = \begin{cases} 
K & \text{if } c_p \leq c_q \quad \forall c_q \in U \\
0 & \text{otherwise}
\end{cases}
\]

- \( K \) is a normalizing constant so that \( \sum_{p \in U} Pr(p) = 1. \)
- \( K = 1/S \), where \( S \) is the number of shortest paths between O and D.
- Some methods select one shortest path \( p^* \)

\[
Pr(p) = \begin{cases} 
1 & \text{if } p = p^* \\
0 & \text{otherwise}
\end{cases}
\]
Shortest path

Advantages:
- well defined
- no need for behavioral data
- efficient algorithms (Dijkstra)

Disadvantages
- behaviorally unrealistic
- instability with respect to variations in cost
- calibration on real data is very difficult
  - inverse shortest path problem is NP complete

Dial’s approach


**Decision-makers**  all identical  
**Alternatives**  efficient paths between O and D  
**Attributes**  link-additive generalized cost  
**Decision-rules**  multinominal logit model
Dial’s approach

- Def 1: A path is efficient if every link in it has
  - its initial node closer to the origin than its final node, and
  - its final node closer to the destination than its initial node.
- Def 2: A path is efficient if every link in it has its initial node closer to the origin than its final node.

Efficient path: a path that does not backtrack.
Dial’s approach

- Choice set $C_n = \text{set of efficient paths (finite, no loop)}$
- No explicit enumeration
- Every efficient path has a non zero probability to be selected
- Probability to select a path

$$\Pr(p) = \frac{e^{\theta \left( \sum_{(i,j) \in p^*} c(i,j) - \sum_{(i,j) \in p} c(i,j) \right)}}{\sum_{q \in C_n} e^{\theta \left( \sum_{(i,j) \in p^*} c(i,j) - \sum_{(i,j) \in p} q(i,j) \right)}}$$

where $p^*$ is the shortest path and $\theta$ is a parameter
Dial’s approach

Note: the length of the shortest path is constant across $C_n$

\[
Pr(p) = \frac{e^{-\theta \sum_{(i,j) \in p} c(i,j)}}{\sum_{q \in C_n} e^{-\theta \sum_{(i,j) \in q} q(i,j)}} = \frac{e^{-\theta c_p}}{\sum_{q \in C_n} e^{-\theta c_q}}
\]

Multinomial logit model with

\[
V_p = -\theta c_p
\]
Dial’s approach

Advantages:

- probabilistic model, more stable
- calibration parameter $\theta$
- avoid path enumeration
- designed for traffic assignment

Disadvantages:

- MNL assumes independence among alternatives
- efficient paths are mathematically convenient but not behaviorally motivated
Dial’s approach

Path 1: $c$

$$\Pr(1) = \frac{e^{-\theta c_1}}{\sum_{q \in C} e^{-\theta c_q}} \quad \frac{e^{-\theta c}}{3e^{-\theta c}} = \frac{1}{3} \quad \text{for any } c, \delta, \theta$$
Path Size Logit

- With MNL, the utility of overlapping paths is overestimated
- When $\delta$ is large, there is some sort of “double counting”
- Idea: include a correction

$$V_p = -\theta c_p + \beta \ln PS_p$$

where

$$PS_p = \sum_{(i,j) \in p} \frac{c(i,j)}{c_p} \frac{1}{\sum_{q \in C} \delta^q_{i,j}}$$

and

$$\delta^q_{i,j} = \begin{cases} 
1 & \text{if link } (i,j) \text{ belongs to path } q \\
0 & \text{otherwise} 
\end{cases}$$
Path Size Logit

Path 1: \( c \)

\[
\begin{align*}
\text{PS}_1 &= \frac{c}{c} \frac{1}{1} = 1 \\
\text{PS}_2 = \text{PS}_3 &= \frac{c - \delta}{c} \frac{1}{2} + \frac{\delta}{c} \frac{1}{1} = \frac{1}{2} + \frac{\delta}{2c}
\end{align*}
\]
Path Size Logit

\[ \beta = 0.721427 \]
Path Size Logit

\[ \begin{align*}
P(1) &= 0.3 \\
&= 0.35 \\
&= 0.4 \\
&= 0.45 \\
&= 0.5 \\
&= 0.55 \\
&= 0.6 \\
&= 0.65 \\
\end{align*} \]

Modeling transportation networks – p.17/59
Path Size Logit

Advantages:
- MNL formulation: simple
- Easy to compute
- Exploits the network topology
- Practical

Disadvantages:
- Derived from the theory on nested logit
- Several formulations have been proposed
- Correlated with observed and unobserved attributes
- May give biased estimates
Path Size Logit: readings


Path Size Logit: readings


Random utility models

Decision-makers with characteristics
- value of time
- access to information
- trip purpose

Alternatives explicit set of paths

Attributes both link-additive and path specific

Decision-rules RUM designed to capture correlations

Note: MNL is a random utility model, but the independence assumption is inappropriate. We must relax it.
Random utility models

Decision-makers with characteristics
  • value of time
  • access to information
  • trip purpose

Alternatives explicit set of paths

Attributes both link-additive and path specific

Decision-rules RUM designed to capture correlations

Note: MNL is a random utility model, but the independence assumption is inappropriate. We must relax it.

In this lecture, we focus on one of the most complicated issues...
Relax the independence assumption

\[
\begin{pmatrix}
U_{1n} \\
\vdots \\
U_{Jn}
\end{pmatrix} = 
\begin{pmatrix}
V_{1n} \\
\vdots \\
V_{Jn}
\end{pmatrix} + 
\begin{pmatrix}
\varepsilon_{1n} \\
\vdots \\
\varepsilon_{Jn}
\end{pmatrix}
\]

that is

\[U_n = V_n + \varepsilon_n\]

and \(\varepsilon_n\) is a vector of random variables.

Assumption about the random term: multivariate distribution

In the rest, we omit \(n\)
Relax the independence assumption

A multivariate random variable $\varepsilon$ is represented by a density function

$$f(\varepsilon_1, \ldots, \varepsilon_J)$$

and

$$P(\varepsilon \leq x) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_J} f(\varepsilon) \, d\varepsilon_J \cdots d\varepsilon_1$$

where $x \in \mathbb{R}^J$ is a $J \times 1$ vector of constants.
Multinomial probit model

\[
U = \begin{pmatrix}
U_1 \\
\vdots \\
U_J
\end{pmatrix} = \begin{pmatrix}
V_1 + \varepsilon_1 \\
\vdots \\
V_J + \varepsilon_J
\end{pmatrix} = V + \varepsilon
\]

Probability to choose path \( p \)

\[
\Pr(p|C) = \Pr(U_j - U_p \leq 0 \ \forall j \in C)
\]

Let \( \Delta_i \) be a \((J - 1 \times J)\) matrix obtained from the \((J - 1 \times J - 1)\) identity matrix, where a column of \(-1\) has been added at index \( i \).

For \( J = 3 \), we have

\[
\begin{align*}
\Delta_1 &= \begin{pmatrix}
-1 & 1 & 0 \\
-1 & 0 & 1
\end{pmatrix} \\
\Delta_2 &= \begin{pmatrix}
1 & -1 & 0 \\
0 & -1 & 1
\end{pmatrix} \\
\Delta_3 &= \begin{pmatrix}
1 & 0 & -1 \\
0 & 1 & -1
\end{pmatrix}
\end{align*}
\]
Multinomial probit model

\[ U = \begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix}, \quad \Delta_2 = \begin{pmatrix} 1 & -1 & 0 \\ 0 & -1 & 1 \end{pmatrix} \]

Therefore

\[ \Delta_2 U = \begin{pmatrix} U_1 - U_2 \\ U_3 - U_2 \end{pmatrix} \]

In general, we obtain

\[
\Pr(p | C) = \Pr(U_j - U_p \leq 0 \ \forall j \in C) = \Pr(\Delta_p U \leq 0)
\]
Multinomial probit model

Assume that \( U \sim N(V, \Sigma) \), where \( \Sigma \in \mathbb{R}^{J \times J} \) is the variance-covariance matrix.

\[
\Delta_p U \sim N(\Delta_p V, \Delta_p \Sigma \Delta_p^T)
\]

\[
\Pr(p|C) = \int_{\varepsilon_J = -\infty}^{0} \ldots \int_{\varepsilon_{p-1} = -\infty}^{0} \int_{\varepsilon_{p+1} = -\infty}^{0} \ldots \int_{\varepsilon_1 = -\infty}^{0} f_p(\varepsilon) d\varepsilon_1 \ldots d\varepsilon_{p-1} d\varepsilon_{p+1} \ldots d\varepsilon_J
\]

\[
f_p(\varepsilon) = (2\pi)^{-\frac{J-1}{2}} |\Delta_p \Sigma \Delta_p^T|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} (\varepsilon - \Delta_p V)^T (\Delta_p \Sigma \Delta_p^T)^{-1} (\varepsilon - \Delta_p V) \right)
\]
Multinomial probit: issues

Variance-covariance matrix

- must be independent of the OD
- must capture the physical overlap of paths
- contains $J(J + 1)/2$ unknown parameters
- Idea: use a structured variance-covariance matrix, with few parameters

Structured variance-covariance

\[ \varepsilon = \varepsilon^l + \varepsilon^0 \]

where \( \varepsilon^0 \) is a vector of independent r.v.
It is assumed that

\[ \text{var}(\varepsilon^l_r) = c_r \sigma^2 \]

where

- \( c_r \) is the length (or travel time) of path \( r \)
- \( \sigma \) is an unknown parameter to be estimated

The covariance is computed as follows:

\[ \text{cov}(\varepsilon^l_r, \varepsilon^l_q) = \mathbb{E}[\varepsilon^l_r \varepsilon^l_q] - \mathbb{E}[\varepsilon^l_r] \mathbb{E}[\varepsilon^l_q] = \mathbb{E}[\varepsilon^l_r \varepsilon^l_q] \]
Structured variance-covariance

Let

$$\varepsilon_r^\ell = \varepsilon_{r \cap q} + \varepsilon_{r \setminus q}$$

where

- $\varepsilon_{r \cap q}$ corresponds to sections of $r$ overlapping with $q$
- $\varepsilon_{r \setminus q}$ corresponds to sections of $r$ not overlapping with $q$

$$E[\varepsilon_r^\ell \varepsilon_q^\ell] = E[(\varepsilon_{r \cap q} + \varepsilon_{r \setminus q})(\varepsilon_{r \cap q} + \varepsilon_{q \setminus r})] = E[\varepsilon_{r \cap q}^2] = \text{var}(\varepsilon_{r \cap q})$$

As before, we define

$$\text{var}(\varepsilon_{r \cap q}) = c_{rq} \sigma^2$$

where $c_{rq}$ is the length (or travel time) of the section common to $r$ and $q$. 
Structured variance-covariance

\[ \varepsilon = \varepsilon^\ell + \varepsilon^0, \quad \Sigma = \Sigma^\ell + \Sigma^0 \]

with

\[ \Sigma^\ell = \sigma^2 \begin{pmatrix} c_1 & c_{12} & \cdots & c_{1J} \\ c_{12} & c_2 & \cdots & c_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ c_{1J} & c_{2J} & \cdots & c_J \end{pmatrix} \]

and

\[ \Sigma^0 = \sigma_0^2 I \]

Note: only the ratio \( \sigma / \sigma_0 \) is identified
Multinomial probit: issues

Estimation

- Integral has no closed form
- Numerical integration is cumbersome and almost impossible, except when \( J \) is small
- **Most efficient estimation method:** Bolduc (1999) A practical technique to estimate multinomial probit models in transportation *Transportation Research Part B* 33, pp. 63-79.
- Bolduc estimates a model with 9 alternatives
- Yai et. al. estimate a model with \( J = 3 \).
- In the route choice context, \( J \) can be much larger
Relax the independence assumption

- If the CDF $F(\varepsilon_1, \ldots, \varepsilon_J)$ of the distribution is known

$$f(\varepsilon_1, \ldots, \varepsilon_J) = \frac{\partial^J F}{\partial \varepsilon_1 \cdots \partial \varepsilon_J}(\varepsilon_1, \ldots, \varepsilon_J)$$

- The choice probability is

$$\Pr(p) = \Pr(V_1 + \varepsilon_1 \leq V_p + \varepsilon_p, \ldots, V_J + \varepsilon_J \leq V_p + \varepsilon_p)$$

$$= P(\varepsilon_1 \leq V_p + \varepsilon_p - V_1, \ldots, \varepsilon_J \leq V_p + \varepsilon_p - V_J)$$

$$= \int_{\varepsilon_p = -\infty}^{\infty} F_p(V_p + \varepsilon_p - V_1, \ldots, \varepsilon_p, \ldots, V_p + \varepsilon_p - V_J) d\varepsilon_p$$
**MEV models**

Family of models proposed by McFadden (1978) (called GEV) Idea: a model is generated by a function

\[ G : \mathbb{R}^J \rightarrow \mathbb{R} \]

From \( G \), we can build

- The cumulative distribution function (CDF)
- The probability model
- The expected maximum utility

MEV models

1. $G$ is homogeneous of degree $\mu > 0$, that is

$$G(\alpha y) = \alpha^\mu G(y)$$

2. $\lim_{y_i \rightarrow +\infty} G(y_1, \ldots, y_i, \ldots, y_J) = +\infty$, for each $i = 1, \ldots, J$,

3. the $k$th partial derivative with respect to $k$ distinct $y_i$ is non negative if $k$ is odd and non positive if $k$ is even, i.e., for all (distinct) indices $i_1, \ldots, i_k \in \{1, \ldots, J\}$, we have

$$(-1)^k \frac{\partial^k G}{\partial y_{i_1} \cdots \partial y_{i_k}}(y) \leq 0, \ \forall y \in \mathbb{R}_+^J.$$
MEV models

- CDF: $F(\varepsilon_1, \ldots, \varepsilon_J) = e^{-G(e^{-\varepsilon_1}, \ldots, e^{-\varepsilon_J})}$

- Probability: $P(i|C) = \frac{e^{V_i + \ln G_i(e^{V_1}, \ldots, e^{V_J})}}{\sum_{j\in C} e^{V_j + \ln G_j(e^{V_1}, \ldots, e^{V_J})}}$ with $G_i = \frac{\partial G}{\partial x_i}$. This is a closed form

- Expected maximum utility: $V_C = \ln G(\ldots) + \gamma$ where $\gamma$ is Euler’s constant.

- Note: $P(i|C) = \frac{\partial V_C}{\partial V_i}$.

Euler’s constant

$$\gamma = -\int_0^{+\infty} e^{-x} \ln x \, dx = \lim_{n \to \infty} \left( \sum_{k=1}^{n} \frac{1}{k} - \ln n \right) \approx 0.57722$$
Network representation of MEV models

Automatic way of defining $G$ based on a network representation

Let $(V, E)$ be a network with link parameters $\alpha_{(i,j)} \geq 0$

Assumptions:

1. No circuit.
2. One node without predecessor: root.
4. For each node $v_i$, there exists at least one path from the root to $v_i$ such that $\prod_{k=1}^{P} \alpha(i_{k-1}, i_k) > 0$. 
Network MEV

For each node $v_i$, we define

- a set of indices $I_i \subseteq \{1, \ldots, J\}$ of $J_i$ relevant alternatives,
- a homogeneous function $G^i : \mathbb{R}^{J_i} \rightarrow \mathbb{R}$, and
- a parameter $\mu_i$.

**Recursive definition of $I_i$:**

- $I_i = \{i\}$ for alternatives,
- $I_i = \bigcup_{j \in \text{succ}(i)} I_j$ for all other nodes.
Network MEV

Recursive definition of $G^i$:

For alternatives:

\[
G^i : \mathbb{R} \rightarrow \mathbb{R} \quad : \quad G^i(x_i) = x_i^{\mu_i} \quad i = 1, \ldots, J
\]

For all others:

\[
G^i : \mathbb{R}^{J_i} \rightarrow \mathbb{R} \quad : \quad G^i(x) = \sum_{j \in \text{succ}(i)} \alpha(i,j) G^j(x)^{\mu_i^{\mu_j}}
\]
Network MEV

Example: Cross-Nested Logit

\[ G = \sum_{m} \left( \sum_{j \in C} \alpha_{jm} y_{j}^{\mu_{m}} \right) \]

\[ \sum_{i=4,5} \alpha_{0i} \left( \alpha_{i1} y_{1}^{\mu_{i}} + \alpha_{i2} y_{2}^{\mu_{i}} + \alpha_{i3} y_{3}^{\mu_{i}} \right) \]

\[ \alpha_{51} y_{1}^{\mu_{5}} + \alpha_{52} y_{2}^{\mu_{5}} + \alpha_{53} y_{3}^{\mu_{5}} \]
Network MEV


Idea:

- Use the cross-nested model specification
- Alternatives = paths
- Nests = links
Link Nested Logit Model

\[
\begin{align*}
\ell_1 & \\
\ell_2 & \\
\ell_3 & \\
\ell_4 & \\
p_1 & \\
p_{23} & \\
p_{24} &
\end{align*}
\]
Link Nested Logit Model

- In this example, nests $\ell_1$, $\ell_3$ and $\ell_4$ contain a single alternative
- The model collapses to a nested logit model
Link Nested Logit Model
Link Nested Logit Model

\[
\begin{align*}
\mu_1 & \rightarrow \alpha_{13} \rightarrow \mu_3 \\
\mu_3 & \rightarrow \alpha_{32} \rightarrow \mu_2 \\
\mu_2 & \rightarrow \alpha_{21} \rightarrow \mu_4 \\
\mu_2 & \rightarrow \alpha_{23} \rightarrow \mu_5 \\
\mu_5 & \rightarrow \alpha_{42} \rightarrow \mu_4
\end{align*}
\]
Link Nested Logit Model

- $5 + 12 = 17$ parameters to capture the correlation
- Variance covariance matrix: $\Sigma \in \mathbb{R}^{4 \times 4}$
- It is symmetric, so there are $J(J + 1)/2 = 10$ entries
- For a given correlation matrix, the identification of the associated parameters is not straightforward

Link Nested Logit Model

Advantages
- Closed form probability
- Rich correlation structure

Disadvantages
- High number of nests
- All correlations structures cannot be reproduced
- Identification of the associated parameters is not straightforward
Factor Analytic Specification


Subnetwork component

Sequence of links corresponding to a part of the network which can be easily labeled, and is behaviorally meaningful in actual route descriptions

- Champs-Elysées in Paris
- Fifth Avenue in New York
- Mass Pike in Boston
- City center in Lausanne

Paths sharing a subnetwork component are assumed to be correlated even if they are not physically overlapping.
Subnetworks - Example
Subnetworks - Methodology

- Factor analytic specification of an error component model (based on model presented in Bekhor et al., 2002)

\[ U_p = V_p + \sum_s \sqrt{c_{ps}} \sigma_s \zeta_s + \nu_p \]

- \( c_{ps} \) is the length by which path \( p \) overlaps with subnetwork component \( s \)
- \( \sigma_s \) is an unknown parameter
- \( \zeta_s \sim N(0, 1) \)
- \( \nu_p \) i.i.d. Extreme Value distributed
Subnetworks - Example

\[ U_1 = \beta^T X_1 + \sqrt{l_{1a}} \sigma_a \zeta_a + \sqrt{l_{1b}} \sigma_b \zeta_b + \nu_1 \]
\[ U_2 = \beta^T X_2 + \sqrt{l_{2a}} \sigma_a \zeta_a + \nu_2 \]
\[ U_3 = \beta^T X_3 + \sqrt{l_{3b}} \sigma_b \zeta_b + \nu_3 \]

\[ \Sigma = \begin{bmatrix} l_{1a} \sigma_a^2 + l_{1b} \sigma_b^2 & \sqrt{l_{1a}} \sqrt{l_{2a}} \sigma_a^2 & \sqrt{l_{1b}} \sqrt{l_{3b}} \sigma_b^2 \\ \sqrt{l_{1a}} \sqrt{l_{2a}} \sigma_a^2 & l_{2a} \sigma_a^2 & 0 \\ \sqrt{l_{3b}} \sqrt{l_{1b}} \sigma_b^2 & 0 & l_{3b} \sigma_b^2 \end{bmatrix} \]
Mixture of MNL

In statistics, a mixture density is a pdf which is a convex linear combination of other pdf’s. If \( f(\varepsilon, \theta) \) is a pdf, and if \( w(\theta) \) is a nonnegative function such that

\[
\int_{\theta} w(\theta) d\theta = 1
\]

then

\[
g(\varepsilon) = \int_{\theta} w(\theta) f(\varepsilon, \theta) d\theta
\]

is also a pdf. We say that \( g \) is a mixture of \( f \). If \( f \) is the pdf of a MNL model, it is a mixture of MNL.
Mixture of MNL

\[ U_p = V_p + \sum_s \sqrt{c_{ps}} \sigma_s \zeta_s + \nu_p \]

If \( \zeta \) is given,

\[ \Pr(p|\zeta) = \frac{e^{V_p + \sum_s \sqrt{c_{ps}} \sigma_s \zeta_s}}{\sum_q e^{V_q + \sum_s \sqrt{c_{qs}} \sigma_s \zeta_s}} \]

\( \zeta \) is distributed \( N(0, I) \)

\[ \Pr(p) = \int \frac{e^{V_p + \sum_s \sqrt{c_{ps}} \sigma_s \zeta_s}}{\sum_q e^{V_q + \sum_s \sqrt{c_{qs}} \sigma_s \zeta_s}} \phi(\zeta) d\zeta \]
Mixture of MNL

\[ Pr(p) = \int \frac{e^{V_p + \sum_s \sqrt{c_{ps}} \sigma_s \zeta_s}}{\sum_q e^{V_q + \sum_s \sqrt{c_{qs}} \sigma_s \zeta_s}} \phi(\zeta) d\zeta \]

Not a closed form. Simulated Maximum Likelihood is to be used

Subnetworks

Advantages

- Rich correlation structure
- Flexibility between complexity and realism

Disadvantages

- Non closed form
Summary

- Shortest path
- Dial’s efficient paths
- Path Size Logit
- Multinomial probit
- Link Nested Logit
- Subnetworks
Additionnal Reading


