Advances in route choice modelling

Michel Bierlaire

TRANSP-OR Transport and Mobility Laboratory
Ecole Polytechnique Fédérale de Lausanne
transp-or.epfl.ch
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
- \( C_n = U \) \( \forall n \)
- \( U \) 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

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

- $K$ 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** multinomial 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: $c$

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

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

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: \( c \)

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

\[ P(1) \]

\[ \delta = 0.721427 \]
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

- **MEV models:**
    doi:10.1016/j.trb.2006.11.006

Relax the independence assumption

- Mixtures of MNL:
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

Path 1  Path 2  Path 3

$S_a$

$S_b$
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

\begin{align*}
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
\end{align*}

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

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

If \( \xi \) is given,

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

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

\[ \Pr(p) = \int_{\xi} \frac{e^{V_p + \sum_s \sqrt{c_{ps}} \sigma_s \xi_s}}{\sum_q e^{V_q + \sum_s \sqrt{c_{qs}} \sigma_s \xi_s}} \phi(\xi) d\xi \]
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
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
Introduction

Set of all paths $\mathcal{U}$ from $o$ to $d$:

- Deterministic
  - $\mathcal{M} \subseteq \mathcal{U}$

- Stochastic
  - $\mathcal{M}_n \subseteq \mathcal{U}$

Path generation:

Route choice Choice set formation:

Route choice model:

$$P(i|C_n) \quad P(i) = \sum_{C_n \in G_n} P(i|C_n)P(C_n)$$
Introduction

- Underlying assumption in existing approaches: the actual choice set is generated
- Empirical results suggest that this is not always true
- Our approach:
  - True choice set = universal set $\mathcal{U}$
  - Too large
  - Sampling of alternatives
Sampling of Alternatives

- Multinomial Logit model (e.g. Ben-Akiva and Lerman, 1985):

\[
P(i|C_n) = \frac{q(C_n|i) P(i)}{\sum_{j \in C_n} q(C_n|j) P(j)} = \frac{e^{V_{in}} + \ln q(C_n|i)}{\sum_{j \in C_n} e^{V_{jn}} + \ln q(C_n|j)}
\]

- \(C_n\): set of sampled alternatives
- \(q(C_n|j)\): probability of sampling \(C_n\) given that \(j\) is the chosen alternative
Importance Sampling of Alternatives

- Attractive paths have higher probability of being sampled than unattractive paths
- Path utilities must be corrected in order to obtain unbiased estimation results
Path Sampling

- Idea: random walk, biased toward the shortest path
- Probability of selecting next link $\ell$ depends on its weight $\omega_\ell$
- Kumaraswamy distribution,

$$\omega(\ell|a, b) = 1 - (1 - x_\ell^a)^b, \quad x_\ell \in [0, 1].$$

$$x_\ell = \frac{SP(v,d)}{C(\ell)+SP(w,d)}$$
Path Sampling

\[ F(x_{\ell} \mid a, b) \]

\[ a = 1, 2, 5, 10, 30 \]
\[ b = 1 \]
Path Sampling

- Probability for path $j$ to be sampled

\[ q(j) = \prod_{\ell = (v, w) \in \Gamma_j} q((v, w) | E_v) \]

- $\Gamma_j$: ordered set of all links in $j$
- $v$: source node of $j$
- $E_v$: set of all outgoing links from $v$
- In theory, the set of all paths $U$ may be unbounded. We treat it as bounded with size $J$
Path Sampling

The correction term can be computed.


\[
P(i|C_n) = \frac{e^{V_{in} + \ln \left( \frac{k_i}{q(i)} \right)}}{\sum_{j \in C_n} e^{V_{jn} + \ln \left( \frac{k_j}{q(j)} \right)}}
\]
Numerical Results

- Estimation of models based on synthetic data generated with a postulated model
- Evaluation of
  - Sampling correction
  - Path Size attribute
  - Biased random walk algorithm parameters
Numerical Results

Diagram of network with nodes and connections labeled as SB.
Numerical Results

- True model: Path Size Logit
  \[ U_j = \beta_{PS} \ln PS_j^U + \beta_L \text{Length}_j + \beta_{SB} \text{SpeedBumps}_j + \varepsilon_j \]
  \[ \beta_{PS} = 1, \quad \beta_L = -0.3, \quad \beta_{SB} = -0.1 \]
  \( \varepsilon_j \) distributed Extreme Value with scale 1 and location 0

- \( PS_j^U = \sum_{\ell \in \Gamma_j} \frac{L_\ell}{L_j} \frac{1}{\sum_{p \in U} \delta_{lp}} \)

- 3000 observations
Numerical Results

- Four model specifications

<table>
<thead>
<tr>
<th></th>
<th>Sampling Correction</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Without</td>
<td>With</td>
</tr>
<tr>
<td>Path Size</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$C$</td>
<td>$M_{PS(C)}^{NoCorr}$</td>
<td>$M_{PS(C)}^{Corr}$</td>
</tr>
<tr>
<td>$U$</td>
<td>$M_{PS(U)}^{NoCorr}$</td>
<td>$M_{PS(U)}^{Corr}$</td>
</tr>
</tbody>
</table>

\[
PS_{i}^{U} = \sum_{\ell \in \Gamma_i} \frac{L_{\ell}}{L_i} \frac{1}{\sum_{j \in U} \delta_{\ell j}} \\
PS_{in}^{C} = \sum_{\ell \in \Gamma_i} \frac{L_{\ell}}{L_i} \frac{1}{\sum_{j \in C_n} \delta_{\ell j}}
\]
Numerical Results

- **Model** $M_{PS(C)}^{NoCorr}$:
  \[ V_{in} = \mu \left( \beta_{PS} \ln PS_{in}^C - 0.3 \text{Length}_i + \beta_{SB} \text{SpeedBumps}_i \right) \]

- **Model** $M_{PS(C)}^{Corr}$:
  \[ V_{in} = \mu \left( \beta_{PS} \ln PS_{in}^C - 0.3 \text{Length}_i + \beta_{SB} \text{SpeedBumps}_i + \ln \left( \frac{k_i}{q(i)} \right) \right) \]

- **Model** $M_{PS(U)}^{NoCorr}$:
  \[ V_{in} = \mu \left( \beta_{PS} \ln PS_{in}^U - 0.3 \text{Length}_i + \beta_{SB} \text{SpeedBumps}_i \right) \]

- **Model** $M_{PS(U)}^{Corr}$:
  \[ V_{in} = \mu \left( \beta_{PS} \ln PS_{in}^U - 0.3 \text{Length}_i + \beta_{SB} \text{SpeedBumps}_i + \ln \left( \frac{k_i}{q(i)} \right) \right) \]
## Numerical Results

<table>
<thead>
<tr>
<th></th>
<th>True</th>
<th>$M_{PS(C)}^{NoCorr}$</th>
<th>$M_{PS(C)}^{Corr}$</th>
<th>$M_{PS(\mathcal{U})}^{NoCorr}$</th>
<th>$M_{PS(\mathcal{U})}^{Corr}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSL</td>
<td>PSL</td>
<td>PSL</td>
<td>PSL</td>
<td>PSL</td>
</tr>
<tr>
<td>$\widehat{\beta}_L$ fixed</td>
<td>-0.3</td>
<td>-0.3</td>
<td>-0.3</td>
<td>-0.3</td>
<td>-0.3</td>
</tr>
<tr>
<td>$\widehat{\mu}$</td>
<td>1</td>
<td>0.182</td>
<td>0.724</td>
<td>0.141</td>
<td>0.994</td>
</tr>
<tr>
<td>Standard error</td>
<td></td>
<td>0.0277</td>
<td>0.0226</td>
<td>0.0263</td>
<td>0.0286</td>
</tr>
<tr>
<td>t-test w.r.t. 1</td>
<td></td>
<td>-29.54</td>
<td>-12.21</td>
<td>-32.64</td>
<td>-0.2</td>
</tr>
<tr>
<td>$\widehat{\beta}_{PS}$</td>
<td>1</td>
<td>1.94</td>
<td>0.411</td>
<td>-1.02</td>
<td>1.04</td>
</tr>
<tr>
<td>Standard error</td>
<td></td>
<td>0.428</td>
<td>0.104</td>
<td>0.383</td>
<td>0.0474</td>
</tr>
<tr>
<td>t-test w.r.t. 1</td>
<td></td>
<td>2.20</td>
<td>-5.66</td>
<td>-5.27</td>
<td>0.84</td>
</tr>
<tr>
<td>$\widehat{\beta}_{SB}$</td>
<td>-0.1</td>
<td>-1.91</td>
<td>-0.226</td>
<td>-2.82</td>
<td>-0.0867</td>
</tr>
<tr>
<td>Standard error</td>
<td></td>
<td>0.25</td>
<td>0.0355</td>
<td>0.428</td>
<td>0.0238</td>
</tr>
<tr>
<td>t-test w.r.t. -0.1</td>
<td></td>
<td>-7.24</td>
<td>-3.55</td>
<td>-6.36</td>
<td>0.56</td>
</tr>
</tbody>
</table>
## Numerical Results

<table>
<thead>
<tr>
<th></th>
<th>True PSL</th>
<th>$M_{PS(C)}^{NoCorr}$ PSL</th>
<th>$M_{PS(C)}^{Corr}$ PSL</th>
<th>$M_{PS(U)}^{NoCorr}$ PSL</th>
<th>$M_{PS(U)}^{Corr}$ PSL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Final Log-likelihood</td>
<td>-6660.45</td>
<td>-6082.53</td>
<td>-6666.82</td>
<td>-5933.98</td>
<td></td>
</tr>
<tr>
<td>Adj. Rho-square</td>
<td>0.018</td>
<td>0.103</td>
<td>0.017</td>
<td>0.125</td>
<td></td>
</tr>
</tbody>
</table>

Null Log-likelihood: -6784.96, 3000 observations

Algorithm parameters: 10 draws, $a = 5$, $b = 1$, $C(\ell) = L_\ell$

Average size of sampled choice sets: 9.66

BIOGEME (Bierlaire, 2007 and Bierlaire, 2003) has been used for all model estimations
Extended Path Size

- Compute Path Size attribute based on an *extended choice set* $C^\text{extended}_n$
- Simple random draws from $\mathcal{U}\setminus C_n$ so that $|C_n| \leq |C^\text{extended}_n| \leq |\mathcal{U}|$
Extended Path Size

Average number of paths in $C_n^{\text{extended}}$ vs. T-test w.r.t. true value

- Path Size
- Speed Bump
- Scale Parameter
Extended Path Size

- Heuristic for finding an extended choice set $C_n^{\text{extended}}$ (all paths in $C_n$ are included)
- Frejinger and Bierlaire (2007)
## Extended Path Size

<table>
<thead>
<tr>
<th></th>
<th>True PSL</th>
<th>(PS(C^{\text{extended}})) PSL</th>
<th>(PS(C)) PSL</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\hat{\beta}_L) fixed</td>
<td>-0.3</td>
<td>-0.3</td>
<td>-0.3</td>
</tr>
<tr>
<td>(\hat{\mu})</td>
<td>1</td>
<td>0.885</td>
<td>0.724</td>
</tr>
<tr>
<td>Standard error</td>
<td></td>
<td>0.0259</td>
<td>0.0266</td>
</tr>
<tr>
<td>t-test w.r.t. 1</td>
<td></td>
<td>-4.43</td>
<td>-12.21</td>
</tr>
<tr>
<td>(\hat{\beta}_{PS})</td>
<td>1</td>
<td>1.52</td>
<td>0.411</td>
</tr>
<tr>
<td>Standard error</td>
<td></td>
<td>0.102</td>
<td>0.104</td>
</tr>
<tr>
<td>t-test w.r.t. 1</td>
<td></td>
<td>5.10</td>
<td>-5.66</td>
</tr>
<tr>
<td>(\hat{\beta}_{SB})</td>
<td>-0.1</td>
<td>-0.131</td>
<td>-0.266</td>
</tr>
<tr>
<td>Standard error</td>
<td></td>
<td>0.0281</td>
<td>0.0355</td>
</tr>
<tr>
<td>t-test w.r.t. -0.1</td>
<td></td>
<td>-1.10</td>
<td>-3.55</td>
</tr>
<tr>
<td>Adj. Rho-Squared</td>
<td></td>
<td>0.114</td>
<td>0.103</td>
</tr>
<tr>
<td>Final Log-likelihood</td>
<td></td>
<td>-6006.96</td>
<td>-6082.53</td>
</tr>
</tbody>
</table>
Conclusions

- Route choice models complicated because:
  1. Complex correlation structure
  2. Large set of alternatives
- Flexible correlation structure: the subnetwork approach
- New point of view on choice set generation and route choice modeling: the path sampling approach

Thank you to Emma Frejinger